UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

Centro de Ciencias Básicas

TESIS

Algoritmos Inteligentes Para Administración y uso del Tiempo Centrados en la Preservación y Aumento del Bienestar Humano

Intelligent Algorithms for Time Management and use Focused on Preserving and Enhancing Human Wellbeing

PRESENTA

Mario Eduardo Marín Limón

PARA OBTENER EL GRADO DE DOCTOR EN CIENCIAS APLICADAS Y TECNOLOGÍA

TUTORES

Dr. Julio César Ponce Gallegos

Dr. Ángel Eduardo Muñoz Zavala

INTEGRANTE DEL COMITÉ TUTORAL

Dr. Alejandro Padilla Díaz

Aguascalientes, Ags. 29 de noviembre de 2023

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

CARTA DE VOTO APROBATORIO
COMITE TUTORAL

M. en C. Jorge Martín Alférez Chavez
DECANO (A) DEL CENTRO DE CIENCIAS BÁSICAS

P R E S E N T E

Por medio del presente como **Miembros del Comité Tutoral** designado del estudiante **MARIO EDUARDO MARIN LIMON** con ID 279929 quien realizó la tesis titulada: **ALGORITMOS INTELIGENTES PARA ADMINISTRACIÓN Y USO DEL TIEMPO CENTRADOS EN LA PRESERVACIÓN Y AUMENTO DEL BIENESTAR HUMANO**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia damos nuestro consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que nos permitimos emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirlo así como continuar con el procedimiento administrativo para la obtención del grado.

Ponemos lo anterior a su digna consideración y sin otro particular por el momento, le enviamos un cordial saludo.

A T E N T A M E N T E
"Se Lumen Proferre"
Aguascalientes, Ags., a 29 de noviembre de 2023.

Dr. Julio César Ponce Gallegos
Co-Tutor de *tesis*

Dr. Ángel Eduardo Muñoz Zavala
Co-Tutor de *tesis*

Dr. Alejandro Padilla Díaz
Asesor de *tesis*

c.c.p.- Interesado
c.c.p.- Secretaría Técnica del Programa de Posgrado

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

DICTAMEN DE LIBERACION ACADEMICA PARA INICIAR LOS TRAMITES DEL EXAMEN DE GRADO

Fecha de dictaminación dd/mm/aaaa: 29/11/2023

NOMBRE: Mario Eduardo Marín Limón **ID** 279929

PROGRAMA: Doctorado en Ciencias Aplicadas y Tecnología **LGAC (del posgrado):** Inteligencia Artificial y Modelación Estadística

TIPO DE TRABAJO: (X) Tesis ( ) Trabajo Práctico

TITULO: ALGORITMOS INTELIGENTES PARA ADMINISTRACIÓN Y USO DEL TIEMPO CENTRADOS EN LA PRESERVACIÓN Y AUMENTO DEL BIENESTAR HUMANO

IMPACTO SOCIAL (señalar el impacto logrado): La tesis se enfoca en la creación de un algoritmo recomendador que indica al usuario las actividades que debe realizar o dejar de hacer con el objetivo de maximizar su bienestar en el contexto de la calidad de vida.

INDICAR SI NO N.A. (NO APLICA) SEGÚN CORRESPONDA:

	Elementos para la revisión académica del trabajo de tesis o trabajo práctico:
SI	El trabajo es congruente con las LGAC del programa de posgrado
SI	La problemática fue abordada desde un enfoque multidisciplinario
SI	Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI	Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI	Los resultados presentados en el trabajo son de gran relevancia científica, tecnologíca o profesional según el área
SI	El trabajo demuestra más de una aportación original al conocimiento de su área
SI	Las aportaciones responden a los problemas prioritarios del país
NO	Generó transferecia del conocimiento o tecnológica
SI	Cumple con la ética para la investigación (reporte de la herramienta antiplagio)
	El egresado cumple con lo siguiente:
SI	Cumple con lo señalado por el Reglamento General de Docencia
SI	Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias,estancia, predoctoral, etc)
SI	Cuenta con los votos aprobatorios del comité tutoral, en caso de los posgrados profesionales si tiene solo tutorpodrá liberar solo el tutor
N.A.	Cuenta con la carta de satisfacción del Usuario
SI	Coincide con el título y objetivo registrado
SI	Tiene congruencia con cuerpos académicos
SI	Tiene el CVU del Conacyt actualizado
SI	Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)
	En caso de Tesis por artículos científicos publicados
N.A.	Aceptación o Publicación de los articulos según el nivel del programa
N.A.	El estudiante es el primer autor
N.A.	El autor de correspondencia es el Tutor del Núcleo Académico Básico
N.A.	En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
N.A.	Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
N.A.	La aceptación o publicación de los artículos en revistas indexadas de alto impacto

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado: **Sí** X **No** ____

FIRMAS

Elaboró:

* NOMBRE Y FIRMA DEL CONSEJERO SEGÚN LA LGAC DE ADSCRIPCION: Dr. Jaime Muñoz Arteaga

NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO: Dr. Angel Eduardo Muñoz Zavala

*** En caso de conflicto de intereses, firmará un revisor miembro del NAB de la LGAC correspondiente distinto al tutor o miembro del comité tutoral, asignado por el Decano**

Revisó:

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO: Dr. Juan Jauregui Rincón

Autorizó:

NOMBRE Y FIRMA DEL DECANO: M.C. Jorge Martín Alférez Chávez

Nota: procede el trámite para el Depto. de Apoyo al Posgrado

En cumplimiento con el Art. 105C del Reglamento General de Docencia que a la letra señala entre las funciones del Consejo Académico: .... Cuidar la eficiencia terminal del programa de posgrado y el Art. 105F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.

Elaborado por: D. Apoyo al Posg.
Revisado por: D. Control Escolar/D. Gestión de Calidad.
Aprobado por: D. Control Escolar/ D. Apoyo al Posg.

Código: DO-SEE-FO-15
Actualización: 01
Emisión: 28/04/20

Article ID: 7250
Subject: Final Verdict
Date: October 11, 2022

XV HIS MICAI 2022

Authors(s): Mario Eduardo Marin Limon, Julio César Ponce Gallegos, Angel Eduardo Muñoz Zavala, Alejandro Padilla Díaz

I hereby write to you to inform you that the work entitled **"Evolutive Algorithms with Restrictions Based in Decision Trees Models for the Enhancement of Satisfaction in Time use"** with **ID 7250**, submitted for the **"XI Workshop of Hybrid Intelligent Systems"** to be held on **October 24-25, 2022** during the MICAI 2022 Workshops day, **was Accepted**, and will be published in a special volume of the journal Research in Computing Science ISSN 1870-4069, in accordance with the following observations:

Reviewer 1 Verdict: Accepted
Recommendations:
The article is well structured and is well supported by the state of the art

Reviewer 2 Verdict: Accepted
Recommendations:
Minor grammar corrections but the paper can be understood

Reviewer 3 Verdict: Accepted
Recommendations:
The article is well supported and with right format

If you have been accepted, we ask you to pay special attention to the dates and requirements published in the page https://hisworkshop.com.mx/

The publication fee amount for accepted papers is $1,500 MXN. Please transfer the fee to any of the following accounts in Mexican pesos to:

Bank: BBVA Bancomer, Account: 0194625285, CLABE: 012180001946252858, Beneficiary: Sociedad Mexicana de Inteligencia Artificial, A.C.

or

Bank: BANAMEX, Branch: 4152, Account: 0194825, CLABE: 002180415201948254, Beneficiary: Sociedad Mexicana de Inteligencia Artificial, A.C.

As payment details please use:

Workshop authors: "MICAI 2022 workshop (workshop abbreviation) paper (paper title or first several words of it)", for example: "MICAI 2022 workshop HIS paper Solving the capacitated vehicle routing problem in a shared bicycle system of a smart city".

RESEARCH IN COMPUTING SCIENCE

ISSN 1870-4069

**Centro de Investigación en Computación, Instituto Politécnico Nacional,
Av. Juan de Dios Bátiz, s/n, Col. La Escalera, CP 07320, DF, México
Tel.: +52-55-5729 6000, ext. 56518, 56653
http://www.rcs.cic.ipn.mx**

Mexico City, July 1st, 2021

To whom it may concern:

Hereby I confirm that the paper

“Peligros y ventajas del diseño de instrumentos de medición auxiliado con estrategias de selección de atributos automáticas”

by Mario Eduardo Marin Limon, Julio César Ponce Gallegos, Angel Eduardo Muñoz Zavala and Alejandro Padilla Díaz

after thorough reviewing process is accepted for publication in our journal.

It is scheduled for the volume 150(5), 2021, which is now in the process of technical production.

With best regards,

…………………………..
Dr. Grigori Sidorov
Editor-in-Chief

Intelligent Time Use Suggestions for Wellbeing Enhancement

Mario E. Marin, Julio C. Ponce

Universidad Autónoma de Aguascalientes,
Centro de Ciencias Básicas,
Mexico

marin@ia.org.mx, jcponce@correo.uaa.mx

Abstract. An intelligent application is developed to automatically suggest to users alternative time assignations to activities, based on data analysis of INEGI's ENUT (national time use survey), with the purpose of enhancing wellbeing. The predictive value of the datasets was ascertained with a comparative analysis of feedforward deep neural networks, support vector machines, logistic regressions and random forest assembles. Logistic regression used the less computational time, and the random forest assembles had the best accuracy. Respondents and users were profiled using K-means clustering, and a non-linear optimization model was developed to find the best datapoints to take suggestions from.

Keywords: Intelligent systems, data analytics, wellbeing, time use, clustering, optimization.

1 Introduction

Time can be considered a scarce resource "whose use largely determines the progress, achievement and wellbeing of individuals, families, communities and societies" [1], therefore managing it to put it to good use should be a priority for individuals and organizations. Despite most time management applications being focused on improving work performance, there is ample evidence that time management practices have a positive correlation with individual wellbeing [2]. However, the research in this field is scarce and distributed among many disciplines [3] and a review of the current applications resulted in very few automated or Artificial Intelligence (AI) approaches, with this field dominated by automated scheduling tools, and some tools for health assessments and wellness. No AI or automated tool was concerned with first suggesting what kind of activities, beyond the preexisting ones, could enhance wellbeing for the user except for tools specifically centered on wellness or health that can hardly take the place of current time management tools in organizations or workplaces where people spend most of their waking hours.

The question this work intended to answer was whether, given an adequate time use dataset, intelligent tools could be developed that automatically consider alternative activities in multiple areas of life from data of similarly positioned individuals in terms of roles and responsibilities, with a wellbeing enhancement criterion.

Agradecimientos

Agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnologías por el apoyo financiero brindado al otorgarme una beca sin la que el presente trabajo no hubiera sido posible. Asimismo, agradezco a la Universidad Autónoma de Aguascalientes por recibirme y por el apoyo económico y de recursos otorgado para que este proyecto llegara a buen puerto. Agradezco también al Gobierno del Estado de Aguascalientes y al Instituto de Ciencia y Tecnología del Estado de Aguascalientes por el apoyo financiero brindado para la investigación aquí desarrollada, así como a la Fundación Telmex-Telcel de la que fui becario para el desarrollo del programa doctoral.

Agradezco a mi comité tutoral. Al Dr. Julio César Ponce Gallegos, quien desde un principio creyó en una investigación que diera prioridad a los criterios de bienestar e impacto social en el contexto del desarrollo de herramientas de Inteligencia Artificial. Su dedicación y paciencia como mentor fue invaluable durante el recorrido que fue el programa doctoral. Agradezco al Dr. Ángel Eduardo Muñoz Zavala y al Dr. Alejandro Padilla Díaz quienes al incorporarse al comité tutoral contribuyeron enormemente con su conocimiento y experiencia en sus respectivas especialidades. Asimismo, agradezco la participación de la Dra. Ma. Antonieta Zuloaga Garmendia y el Dr. Jesús Armando López Velarde Campa en el comité tutoral extendido, en el cual contribuyeron desde sus respectivos enfoques académicos y profesionales a que este trabajo considerara horizontes más amplios.

Agradezco a la comunidad de la Universidad Autónoma de Aguascalientes, a mis compañeros en el doctorado que fueron compañía invaluable, y al personal administrativo y de apoyo que brindaron el soporte necesario para este proyecto fuera posible, así como a la comunidad de investigadores y profesores de la universidad que desinteresadamente contribuyeron con sus opiniones, ideas y retroalimentación a muchas de las ideas aquí presentadas. Agradezco también al Centro de Investigación en Matemáticas y a sus investigadores por su apoyo y compañía en este proyecto.

Agradezco a mi familia que me apoyaron con cariño, entusiasmo y paciencia. En particular gracias a mi padre por nunca estar menos que orgulloso de mí y haberme impulsado con amor y esfuerzo hasta aquí, a mi hermana Sigrid y a mi cuñado Alfredo por su siempre presente apoyo, aliento y amor sin el que no concibo estar donde estoy. Agradezco a mis amigos que alimentaron en mí el deseo y la fortaleza necesaria para emprender este recorrido que es un doctorado. Gracias Benjamín por compartir tu pasión y respeto por la labor académica. Gracias Saúl, Heriberto, Salvador, Juan, Sergio, Jonás y Lorenia por ser inspiraciones con sus respectivos proyectos académicos.

To my father, who spoke into being my desire for academic accomplishment, and to my mother who never settled for just desire and taught me consistent action.

Contents

Table Index

Figure Index

Acronyms

AI	Artificial intelligence
ADHD	Attention deficit hyperactivity disorder
ANN	Artificial neural networks
APA	American Psychological Association
APP	Application (software)
CART	Classification and regression tree
CHAID	Chi-square automatic interaction detection
CNN	Convolutional neural networks
COVID-19	Coronavirus disease 2019
CRM	Customer relationship management
CSS3	Cascading Style Sheets 3
CSV	Comma-separated values
DBSCAN	Density-based spatial clustering of application with noise
DIF	National System for Integral Family Development (Sistema Nacional para el Desarrollo Integral de la Familia)
DL	Deep learning
DNN	Deep neural network
DT	Decision tree
EDA	Electrodermal activity
ENBIARE	National Self-reported Wellbeing Survey (Encuesta Nacional de Bienestar Autorreportado)

ENUT	National Time Use Survey (Encuesta Nacional de uso de Tiempo)
FN	False negative
FP	False positive
GA	Genetic algorithms
GDP	Gross domestic product
GHZ	Gigahertz
GOFAI	Good old-fashioned artificial intelligence
GRU	Gated recurrent units
HTML5	Hypertext Markup Language 5
ID	Identification
ID3	Iterative Dichotomiser 3
INEGI	National Institute for Statistics and Geography (Instituto Nacional de Estadística y Geografía)
INMUJERES	National Institute for Women (Instituto Nacional de las Mujeres)
KNN	K nearest neighbors
LASSO	Least absolute shrinkage and selection operator
LGPDPPSO	General Law on Protection of Personal Data in Possession of Obligated Subjects (Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados)
LLM	Large language model
LR	Logistic regression
LS	Least squares
LSTM	Long short-term memory networks

MIT	Massachusetts Institute of Technology
ML	Machine learning
MLR	Multinomial logistic regression
MXN	Mexican peso(s)
NOM	Official Mexican Norm (Norma Oficial Mexicana)
OCB	Organizational citizen behavior
OECD	Organization for Economic Co-operation and Development
OR	Operations research
PCA	Principal component analysis
PHP	PHP Hypertext Preprocessor
RAM	Random access memory
RF	Random forest
RNN	Recurrent neural networks
SSL	Secure socket layer
SVD	Singular value decompensation
SVM	Support vector machines
SWB	Subjective Wellbeing
TDIDT	Top-down induction of decision trees
TN	True negative
TP	True positive
UAA	Aguascalientes State Autonomous University (Universidad Autónoma de Aguascalientes)

UN	United Nations
USD	United States of America dollars
WAIT	Wellbeing and Agency Intelligent Time Management System
WHO	World Health Organization

Resumen

Esta tesis presenta la implementación de una herramienta inteligente de administración del tiempo, probada en un experimento con estudiantes universitarios, con el objetivo de aportar evidencia de que las implementaciones de algoritmos de inteligencia artificial son una forma eficaz de generar herramientas inteligentes de administración del tiempo significativamente exactas, que podrían ayudar en la toma de decisiones involucradas en el cambio de comportamiento en grupos de personas e individuos para aumentar el bienestar personal y al mismo tiempo equilibrarlo con objetivos organizacionales y sociales, como realizar un conjunto de actividades designadas.

Se introducen conceptos y teorías como bienestar, bienestar subjetivo y administración del tiempo y se conectan con el constructo de bienestar basado en agencia propuesto por el equipo de investigación, así como con la práctica de administración inteligente del tiempo que se propone y explica. Esto con la intención de tener bases sólidas para los criterios utilizados para el diseño de herramientas de administración inteligente del tiempo.

Se presentan los algoritmos inteligentes implementados; para la clasificación se utilizan redes neuronales artificiales profundas, máquinas de soporte vectorial, regresión logística y clasificadores basados en árboles de decisión. Para la búsqueda se utilizan algoritmos genéticos. Todo esto se presenta en el contexto de regímenes de validación cruzada para probar las métricas de rendimiento de los algoritmos. Los conjuntos de datos de entrenamiento para alimentar los algoritmos se obtienen mediante el preprocesamiento de los microdatos de la Encuesta Nacional de Uso del Tiempo del INEGI en sus ediciones de 2014 y 2019. Los conjuntos de datos se caracterizan y se describe el preprocesamiento para facilitar la reproducibilidad de las pruebas presentadas.

Se describe la herramienta de administración inteligente del tiempo desarrollada, incluidos sus resultados de análisis de poder predictivo, su proceso asistido por inteligencia artificial para el diseño de instrumentos de medición, el uso de clasificadores de aprendizaje automático interpretables, el uso de los parámetros de dichos clasificadores para definir un espacio de búsqueda para el algoritmo genético para encontrar soluciones que mejoren el bienestar de los usuarios, así como la base de datos y la interfaz de usuario utilizadas. También se presenta el diseño experimental para la prueba de la herramienta de administración inteligente del tiempo.

Finalmente, los resultados se presentan en forma de resultados de encuestas, validaciones de exactitud, recomendaciones generadas por la herramienta de administración inteligente del tiempo

para los estudiantes y correlación de una variable de bienestar basada en agencia propuesta con otras variables de bienestar. Se discuten los resultados y se llega a la conclusión de que una herramienta inteligente de administración del tiempo, desarrollada con la ayuda de la inteligencia artificial mediante el uso de un proceso de selección de modelos automatizado, puede preservar o mejorar la exactitud mostrada al probar los conjuntos de datos de entrenamiento con clasificadores de aprendizaje máquina en un régimen de validación cruzada, que a su vez tiene mejor exactitud que un clasificador aleatorio con distribución uniforme. Esta exactitud puede ser en una variable o conjunto de variables de interés, como en el caso de la satisfacción con el uso del tiempo en actividades académicas y otras variables de bienestar subjetivo que se probaron en el experimento. Finalmente, se destacan otras contribuciones de la investigación presentada, como el proceso asistido por inteligencia artificial para el diseño de instrumentos de medición que puede usarse en otros estudios de ciencias sociales, y lo prometedor de seguir desarrollando un constructo de bienestar basado en agencia, así como las posibilidades para futuras investigaciones.

Abstract

This dissertation presents the implementation of an intelligent time management tool, that is tested in an experiment with college students, with the objective of contributing evidence that implementations of artificial intelligence algorithms are an effective way to generate meaningfully accurate intelligent time management tools, that could aid in the decision-making involved in changing behavior in groups of people and individuals to increase personal wellbeing while balancing it with organizational and social objectives, such as performing a set of designated activities.

Concepts and theories such as wellbeing, subjective wellbeing, and time management are introduced and connected to the agency-based construct of wellbeing posited by the research team and the practice of intelligent time management that is proposed and explained, with the intention of having solid foundations for the criteria used for the intelligent time management tool design.

The intelligent algorithms implemented are presented. Deep artificial neural networks, support vector machines, logistic regression, and decision tree-based classifiers are used for classification. For search, genetic algorithms are used. All of this is presented in the context of cross-validation regimes to test algorithms' performance metrics. Clean training datasets to feed the algorithms are obtained through preprocessing the microdata of INEGI's Time Use National Survey in its 2014 and 2019 editions. The datasets are characterized, and the preprocessing is described to facilitate the tests' reproducibility.

The intelligent time management tool developed is described, including its predictive power analysis results, its artificial intelligence-aided process for measurement instrument design, the use of interpretable machine learning classifiers, the use of the parameters of such classifiers to define a search space for the genetic algorithm to find solutions that enhance wellbeing for the users, and the database and user interface employed. The experimental design for the intelligent time management tool test is also presented.

Finally, the results are presented in the form of survey results, accuracy validations, recommendations generated by the intelligent time management tool to the students, and correlation of a proposed agency-based wellbeing variable to other wellbeing variables. The results are discussed, and they lead to the conclusion that an intelligent time management tool, developed with the aid of artificial intelligence by using an automated model selection process, can preserve

or enhance the accuracy of testing the training datasets in machine learning classifiers in a cross-validation regime that has better accuracy than a random classifier with a uniform distribution, for a variable or set of variables of interest, as in the case of time-use satisfaction in academic activities and other subjective wellbeing variables that were tested in the experiment. Other contributions of the dissertation are highlighted, such as the artificial intelligence-aided process for measurement instrument design that can be used in other social sciences studies, the promise for further developing an agency-based wellbeing construct, and the possibilities for future research.

I. Introduction

What enhances our ability to act in the world leads to joy, and what decreases this ability leads to sadness. This affirmation is at the core of Spinoza's understanding of the importance of the human capacity to act intentionally in the world, as found in his Ethics (Spinoza, 1996). It is also the thread that runs through this dissertation to connect wellbeing, performance, time management, and artificial intelligence (AI) algorithms to find out if time management tools boosted with such algorithms can predict self-reported wellbeing levels with meaningful accuracy using only the data available to them for their tasks: a small set of personal data from users but primarily their time-use data, in other words, data about action. If these tools can make these predictions, they may help people manage their time while preserving and enhancing their wellbeing. For this purpose, we use state-of-the-art AI algorithms, tests and validations with big datasets, AI-aided measurement instrument design, and robustly defined measurements of subjective wellbeing (SWB) already well linked to overall human wellbeing (Diener & Ryan, 2009), accompanied by data collected based on a wellbeing conceptualization based on human agency, which can be defined as a capacity to act intentionally (Schlosser, 2019), put forward by the research team.

While the benefits of time management practices in enhanced performance and wellbeing are well documented (Claessens et al., 2007), these results show that environmental factors and lack of information limit people's agency, which is an important constraint to these benefits and might even hinder the linkage of time management and SWB itself (Aeon & Aguinis, 2017). In terms of intelligent tools for time management in a data driven environment this means that the accuracy of the algorithms, dependent on the data available, is paramount. Therefore, this dissertation incorporates publicly available big time-use datasets that include SWB data to be used as training datasets, and a dataset collected by the research team with time use and SWB data that also includes data about user agency to be used as test dataset in a data-driven study aided by interpretable classifier algorithms to conduct an experiment to see if an intelligent time management tool can accurately predict SWB levels, which is the main research question.

This dissertation presents both literature findings and contributions by the research team. For instance, literature is presented linking SWB to people's agency without stepping out of the wellbeing paradigm, as well as a definition of agency-based wellbeing proposed by the research team. Also, we present the work done to apply intelligent algorithms that work transparently to aid in decision-making relative to allocating time to different activities in a way that preserves or enhances personal wellbeing, as well as a definition of intelligent time management developed by the research team.

The research presented in this text also incorporates some of the work done to overcome problems of numerical and statistical analysis, measurement instrument design and application, and interpretability for non-technical audiences that arose in the process of finding answers to the research question. Finally, the results are presented and discussed, and some conclusions are drawn regarding their actionable implications and the need for further research.

I.1. Social Impact

The preservation and enhancement of wellbeing can help with public health issues afflicting large parts of the global population. For instance, low levels of SWB are linked to issues such as depression (Cummins, 2013; Gargiulo & Stokes, 2009), addictive behavior (Charzyńska et al., 2021; Koç, 2017), and lower physical health levels (Diener et al., 2017; Okun et al., 1984; Steptoe et al., 2015). SWB levels are also linked to the performance of employees (Bryson et al., 2017; Salgado et al., 2019) and students (Bücker et al., 2018). Combined with time management practices, wellbeing enhancement could potentially be done while aiding people to accomplish in a better way their personal goals and those in the context of the organization where they work or study (Aeon & Aguinis, 2017; Claessens et al., 2007), furthering not only human development but also economic advancement (Durand, 2015).

As inferred by the wealth of evidence referenced in the paragraph above, the social impact of wellbeing preservation and enhancement is well documented. Most prominently, the Organization of Economic Co-operation and Development (OECD) has been measuring and highlighting the importance of wellbeing national metrics, including subjective ones, for over two decades (OECD, 2011b, 2019), stressing that such metrics rival in importance economic ones such as the gross domestic product (GDP) for the development of nations and regions (Nozal, 2019). Wellbeing is thus eminently a public policy issue but, in its implementation, not the exclusive domain of governments as it is understood that the purveyors and implementers of public policy are both governments and organizations such as businesses, corporations, schools, and universities.

I.2. Justification

While efforts to preserve and enhance human wellbeing could be justified as a societal moral imperative, the importance of such efforts and of measuring their impact has been highlighted by the work of governments (Meredith, 2022; Wellbeing Economy Alliance, 2021), and supranational organizations (OECD, 2019; World Health Organization, 2021, 2023). Not only are electorates becoming more interested in wellbeing metrics and demanding their furtherance to their governments (F. Harvey, 2020; Herrin et al., 2018; Park & Peterson, 2019), but also shareholders and other corporative stakeholders have developed a legitimate interest in the risks associated with social issues such as those that cause a lack of preservation of human wellbeing in and by corporations (Baid & Jayaraman, 2022; Becchetti et al., 2022; Bradley, 2021). This is unsurprising given that wellbeing is a reality that traverses all the important dimensions of human life, including health, social connection, financial stability, and safety, among many others that help live a life of happiness and accomplishment (OECD, 2019), and is thus embedded in all human activity.

The wellbeing metrics offered by the OECD (Durand, 2015) highlight the strong environmental components of wellbeing as opposed to it being a reality subject only to individual action and shed light on the social component that any tool to preserve or enhance wellbeing must have. Thus, in the pursuit of social impact, we propose to apply a highly documented social practice such as time management (Aeon & Aguinis, 2017; Claessens et al., 2007) boosted with the power of interpretable machine learning (ML) algorithms that allow their parameters to be inspected and used for further analysis and permit the explicit inspection of their recommendations and predictions while being bound by strict human wellbeing criteria.

I.3. Objectives, hypothesis, and research questions

The general objective of this dissertation is to contribute evidence that implementations of AI algorithms are an effective way to generate meaningfully accurate intelligent time management tools, that could aid in the decision-making involved in changing behavior in groups of people and individuals to increase personal wellbeing while balancing it with organizational and social objectives, such as performing a set of designated activities.

The specific objectives are:

- Articulation of basic principles to increase or conserve human wellbeing as a key value in implementing AI tools.
- Characterization of time-use survey data as human behavior in a time economy where human wellbeing is a key value.
- Implementation of methods for cleaning and processing time use data sets that can be used in intelligent data management tools designed to take human wellbeing into account.
- Implementation of interpretable machine learning algorithms that allow for an easy audit of its parameters, predictions, and recommendations.
- Experiment with a web app time management tool that takes increasing or preserving human well-being as a key value.

The hypothesis for the experiment that pertains to this dissertation is that using AI time management tools trained with mostly time use data (the relevant data to time management) and small set of personal data such as gender, age, and marital status, can predict with meaningful accuracy SWB levels of a group of people not included in the training datasets. Therefore, our null hypotheses need to be worded in terms of an intelligent time management tool with a newly collected test dataset being unable to predict SWB ordinal variables better than with the use of a cross-validation regime using only the training dataset with no test dataset, and this cross-validation regime test having no better accuracy than a random classifier with a uniform distribution. These are the null hypotheses that the data must help reject for the evidence we are looking for to be produced.

Given the regular time structure of college students' activities and the high performance of the AI algorithms used to predict satisfaction with the allocation of time to academic activities (Marin & Ponce, 2020), an experiment in this domain is presented in which the SWB variable of interest is time-use satisfaction in academic activities. Therefore, the main research question is:

- In a sample of college students, can an intelligent time management tool trained with mostly time-use datasets and tested with an experimental dataset have significantly better accuracy predicting the level of time-use satisfaction in academic activities than a random classifier with a uniform distribution, and a cross-validation regime using only the training dataset.

Additionally, the same main research question, and its corresponding hypothesis, can be reformulated with other variables of time-use satisfaction linked to activities other than academic ones, even as the previous research question will continue being the main research question with its focus on academic activities as the experiment will go forward with a sample containing only college students, and a measurement instrument specifically designed to contain questions relevant to time-use satisfaction in academic activities.

Other complementary research questions related to the specific objectives are:

- Can basic principles of preserving and enhancing wellbeing be articulated to develop AI tools?
- Can time-use survey data be characterized as human behavior in an economy of time?
- Can interpretable machine learning algorithms trained mainly with time-use data be used for prediction and recommendations in wellbeing?

I.4. Methodology, research type, and process scheme

To answer the main research question, the experiment uses quantitative methodology using experimental data. However, the AI tools developed for this experiment were tested using a cross-validation regime and trained using existing compiled data from publicly available big datasets. Thus, the first part of the research is done in an exploratory fashion.

The publicly available datasets came from the 2014 and 2019 editions of the National Time Use Survey from the National Institute of Statistics and Geography (INEGI) in Mexico. On the other hand, the experimental data was obtained by a survey expressly designed for this experiment as a part of an intelligent time management tool. AI algorithms aided the measurement instrument design process for the survey by using an automatic model selection strategy. This survey was applied to students of the Basic Sciences Centre of the Aguascalientes State Autonomous University (UAA). The methodological and statistical details of both kinds of datasets are described in full in Chapter V of this dissertation.

The research process was thus composed of three parts. First, a literature review of relevant topics and potentially useful AI algorithms, followed by a search for appropriate datasets to explore these topics, and then the training and testing of the AI algorithms. After these steps, the algorithms that proved helpful in answering the research question were used to design an experiment to test the hypothesis. A scheme of this process is shown in **Figure 1**.

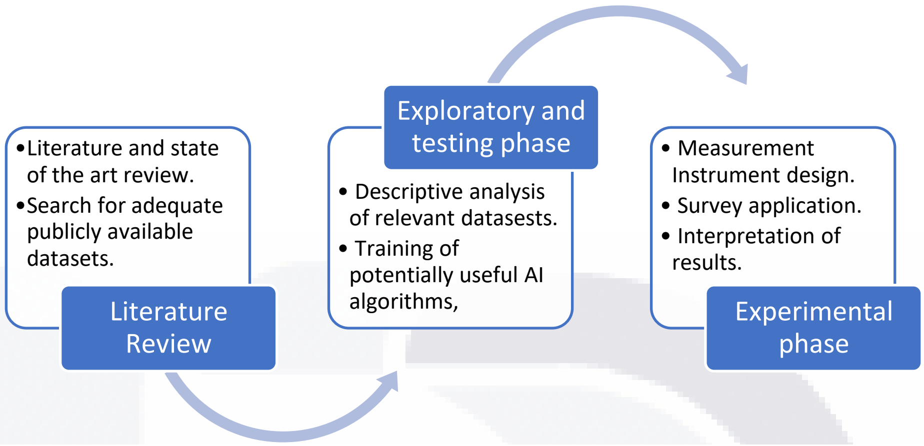


Figure 1: *Process scheme of the research for this dissertation*

I.5. Dissertation's structure

This dissertation is divided into eleven sections that are explained in their contents below:

I. **Introduction:** This is the current section, and it is the first chapter in which the dissertation themes, objectives, justification, and social impact are discussed briefly. Also, the methodology, type of research, and process scheme are presented.

II. **Theoretical Framework:** This section is a chapter that introduces the working definition of wellbeing presented and used by the research team, agency-based wellbeing, and wellbeing constructs compatible with it, such as SWB. The concepts of time use, time management and the basics of the proposed practice of intelligent time management are also introduced.

III. **Literature revision and state of the art:** This section includes the chapter with the results of a literature and state-of-the-art review on the topics related to research and applications on the intersections of time management, wellbeing, and performance. Results of statistical significance known in the domain literature and examples of time management applications that incorporate some aspects of AI are presented.

IV. **Intelligent algorithms used:** In the chapter pertaining to this section, the intelligent algorithms used for classification and search are presented. Artificial neural networks, support vector machines, decision tree-based methods, and logistic regression are presented for classification. For search, genetic algorithms are introduced. Also, deep learning and cross-validation approaches are presented as they are relevant throughout the dissertation.

V. **Datasets:** In the chapter on this section, the two publicly available datasets used for the research presented in the dissertation are described and characterized: the microdata of the Time Use National Survey done in Mexico by INEGI in its 2014 and 2019 editions. Also, the preprocessing to which these datasets were subjected is presented.

VI. **Development of intelligent time management tool:** The intelligent time management tools developed using the intelligent algorithms previously presented are described in the chapter on this section, such as the different classifiers, their predictive power analysis, AI-aided measurement instrument design, genetic algorithm subjective wellbeing enhancement, and the Interpretable machine learning approach used with these tools is presented.

VII. **Experimental design and implementation:** The hypotheses for the experiment are defined and presented, as well as the way in which they will be tested. The practical issues of the experiment are presented such as the sample, survey application, methodology, calendar, and statistical tests are also introduced.

VIII. **Results and discussion:** The experiment's results are fully presented and discussed in the context of the hypothesis and theories already offered. These results include those from the survey, accuracy validation, and the recommendations generated by the intelligent time management tool to the students, as well as the correlation of a proposed agency-based wellbeing variable to other wellbeing variables.

IX. **Conclusions and future research:** This section is the final chapter in which the conclusions and contributions of the dissertation are described, in particular the answers to the research questions, the implications of the results for AI-aided measurement instrument design, the promises of an agency-based wellbeing construct and the answer to the research questions. Also, topics of future research and closing thoughts are offered.

X. **References:** The bibliographic references consulted for this dissertation are listed in this section. The format used for references is the American Psychological Association (APA) 7th edition.

XI. **Annexes:** This section contains all the material that supplements the dissertation such as surveys, messages sent to users, and statistical tables.

II. Theoretical Framework

In this chapter, the theoretical underpinnings of the research done for this dissertation are presented. Definitions, current research, and relevant results about human wellbeing, time use, and time management are presented. These will be used and leaned on throughout the rest of the chapters of this text. Also, the concept of intelligent time management, the practice we are comparing to existing time management practices, is presented and developed, as well as a working definition of agency-based wellbeing posited by the research team.

II.1. Wellbeing

In this subsection, we present a definition of wellbeing based on agency and actions and how these actions relate to what is good for human beings according to their nature, an area of wellbeing studies and philosophy. This emphasis on actions is essential because actions and their consequences can be measured; both can be used to generate data, and such data, when subjected to an analytics process, can result in insights about the impact of such actions on human wellbeing and performance. On the other hand, the emphasis on human nature is a constraint on the types of data collected so that the systems powered by the analytics process primarily serve their users' wellbeing.

It is essential to clarify the definition of wellbeing used in this work because there are many definitions in the literature, some of which have different implications. Some of these definitions can be complementary or compatible in practice with the one we will be presenting, and those are identified. Clarity about what wellbeing is will help us decide what data to use, how to transform and prepare it, and what kind of analysis will be done, including whether and how to train intelligent tools to help individuals and organizations enhance their wellbeing. In other words, clarity about the definition of wellbeing is the first step to defining the value sought in a process that will use AI.

However, what can be considered human nature, and why should we care about it in the context of AI? What are the implications of defining wellbeing in terms of actions? To answer these questions, we need to review some philosophical antecedents of wellbeing that have been used to construct definitions already found in the literature. We will delve deep first to consider the philosophical underpinnings of the concepts behind the words "wellbeing" and "human nature" because we argue that it is crucial to construct a definition that, while compatible with what already exists in the literature, can be more amenable to a high-value data analytics approach in organizations ready to use AI in this context.

II.1.1. Philosophical underpinnings of human wellbeing

For the basics of the modern concepts of wellbeing, Crisp tells us that wellbeing "is most commonly used in philosophy to describe what is non-instrumentally or ultimately good *for* a person […] It has become standard to distinguish theories of well-being as either hedonist theories, desire theories, or objective list theories […] Also important in ethics is the question of how a person's moral character and actions relate to their well-being." (Crisp, 2017).

Even everyday use of the concept of wellbeing implies what is suitable for a person. The question of what is good for a person is of great importance to moral philosophy and Ethics, and there are still many debates open in these areas and their relationship to wellbeing. These debates are of great importance to individuals and organizations, as we will see later; the tools proposed in this work can help researchers of these areas shed light on these questions. For this work, we will take an approach called "perfectionism," which will be explained in the following paragraphs. Let us see the three kinds of theories of wellbeing presented by Crisp (2017).

> ***Hedonist theories****: In answer to the question, 'What does well-being consist in?', then, the hedonist will answer, 'The greatest balance of pleasure over pain'. We might call this substantive hedonism. A complete hedonist position will involve also explanatory hedonism, which consists in an answer to the following question: 'What makes pleasure good, and pain bad?', that answer being, 'The pleasantness of pleasure, and the painfulness of pain'. Consider a substantive hedonist who believed that what makes pleasure good for us is that it fulfills our nature. This theorist is not an explanatory hedonist.* (Crisp, 2017)

While pleasure and pain are undoubtedly important indicators to evaluate wellbeing, particularly in the realm of bodily wellbeing and health, but also in the case of emotional wellbeing, they are not complete indicators of wellbeing. One can quickly think of instances in which optimizing for avoiding pain altogether and encountering only pleasure can go wrong, as in the case of explanatory hedonism in action. However, a balanced view of pain and pleasure is essential to wellbeing.

> ***Desire theories:*** *The simplest version of a desire theory one might call the present desire theory, according to which someone is made better off to the extent that their current*

desires are fulfilled […]. [In] a comprehensive desire theory […] what matters to a person's well-being is the overall level of desire-satisfaction in their life as a whole. A summative version of this theory suggests, straightforwardly enough, that the more desire-fulfilment in a life the better. […] A global version of the comprehensive theory ranks desires, so that desires about the shape and content of one's life as a whole are given some priority […]. According to the informed desire account, the best life is the one I would desire if I were fully informed about all the (non-evaluative) facts. (Crisp, 2017)

Desire theories can lead to a more comprehensive encompassment of what things are part of a life of wellbeing. Nevertheless, these theories also run into the problem that, beyond basic goods like food and shelter, many desires that people have, arise because of their education and interactions with the society they live in and even organizations they belong to, and not because of a need woven into human nature. From this viewpoint, it would be more interesting to analyze data produced by these societies and organizations and compare them to the desires typical of their members. This viewpoint is compatible with social constructivism (Detel, 2001).

Objective list theories: *Usually understood as theories which list items constituting well-being that consist neither merely in pleasurable experience nor in desire-satisfaction. Such items might include, for example, knowledge or friendship. But it is worth remembering, for example, that hedonism might be seen as one kind of 'list' theory, and all list theories might then be opposed to desire theories as a whole. What should go on the list? It is important that every good should be included.*

One common objection to objective list theories is that they are élitist, since they appear to be claiming that certain things are good for people, even if those people will not enjoy them, and do not even want them. (Crisp, 2017)

This study uses an objective list theory approach with a particular explicit criterion for inclusion. We do this partly because modern definitions of wellbeing used by organizations, governments, and supranational entities have been devised as multidimensional constructs like the OECD Well-being Framework (OECD, 2020) or the *Gallup Wellbeing Five* (Healthways, 2015), which are objective list theories of wellbeing. Also, we prefer this approach because objective list theories can include subjective and objective components of wellbeing; as the importance of subjective components is paramount for this work. However, we shall see that even components of wellbeing judged as objective are bound to be subjectively evaluated, according to personal, organizational, and societal factors, about whether they are objective goods and indicators of wellbeing. In other words, in every objective list theory of wellbeing, there is a criterion for inclusion, even if implicitly, for deciding what is a good and, therefore, what should be included. In our case, this criterion is based on the perfectionism approach. Crisp (2017) presents perfectionism thusly:

> *What is the 'good-maker', according to objective list theorists? This depends on the theory. One, influenced by Aristotle and recently developed by Thomas Hurka (1993), is perfectionism, according to which what makes things constituents of well-being is their perfecting human nature. If it is part of human nature to acquire knowledge, for example, then a perfectionist should claim that knowledge is a constituent of well-being. But there is nothing to prevent an objective list theorist's claiming that all that the items on her list have in common is that each, in its own way, advances well-being.* (Crisp, 2017)

Now the question arises. What things perfect human nature? We should first define human nature to see what things could perfect it. However, the definition of human nature is not a settled issue by any means. Not in Philosophy or social sciences. All that we have are working definitions and reflective judgment, as Crisp (2017) puts it: intuition.

> *How do we decide what goes on the list? All we can work on is the deliverance of reflective judgment—intuition, if you like. But one should not conclude from this that objective list theorists are, because they are intuitionist, less satisfactory than the other two theories. For those theories too can be based only on reflective judgment. Nor should one think that intuitionism rules out argument. Argument is one way to bring people to see the truth.* (Crisp, 2017)

Let us rest the issue of human nature for a while and return to perfectionism. As pointed out, the perfectionist approach is influenced by Aristotle, who had his own ideas about human nature, and has been more recently further developed (Hurka, 1993). However, this view dates with sufficient clarity, for our purposes, at least back to Spinoza. We will emphasize perfectionism because, with some help from Spinoza, which along with other more modern philosophers, form the basis of other modern theories of wellbeing, it will let us bypass the conundrum of "what counts as a good?" and let us apply a single criterion to data about people's actions and its effects on their wellbeing; and also avoid the criticism of elitism aimed at definitions of wellbeing based on the objective list theory, like the expertly made OECD Well-being Framework, and the top-research-based Gallup Wellbeing Five, and other objective list theories of wellbeing used in organizations, while not being incompatible with these theories.

II.1.2. Working definition of wellbeing

From Aristotle, we learn that human beings have two main characteristics that are only theirs and have not been found in other known living beings so far, their ability to act rationally and to use symbolic language. Aristotle considers that human beings are the "zoom politikon," or animals that create a city, that is, that engage in the creation of institutions and spaces in which people can live together in harmony, and the "zoom logikon," or animals that think and reason with language (Aristotle, 2013). These two exclusive characteristics of human beings have been relatively uncontroversial both in Philosophy and Science (Look, 2007). No medium of communication used by animals or plants approaches the flexibility and power of human language, which is different in kind and not merely on degree to those of animals and other known living beings. Because of the use of language, human rationality is also a characteristic wholly different in kind from the mental abilities of animals.

In Spinoza's book Ethics, we see that:

> *The idea that of anything that increases or diminishes, aids or restrains, our Body's power of acting, increases or diminishes, aids or restrains, our Mind's power of thinking.*
>
> *We see, then, that the Mind can undergo great changes, and pass now to a greater, now to a lesser perfection. These passions, indeed, explain to us the affects of Joy and Sadness. By Joy, therefore, I shall understand in what follows that passion by which the Mind passes to a greater perfection. And by Sadness, that passion by which it passes to a lesser perfection.* (Spinoza, 1996)

First, it should be noted that Spinoza affirms the unity of body and mind, saying that both mind and body obey the same natural laws as every other object in the universe. In other words, human beings and their rational minds are not separate from nature simply because they are human or rational, only insofar as the laws of nature treat a different kind of object differently from others because of its different properties and not because of some exceptionality inherent to the object. The emphasis of Spinoza in saying this is in naturalizing the human experience and is a foundation of more modern theories that consider humanity an intelligible phenomenon, not beyond the scope of science and reason. It does not mean, or should not be taken to mean, that we should study and treat things found in nature that would exist whether humans existed or not, the same way we treat social phenomena that only exist because groups of people bring them into existence (organizations, for example). As such, even Spinoza continues to talk of body and mind separately when practical.

Also, according to Spinoza (1996), as there is a unity of body and mind; whatever restrains one restrains the other, and whatever enhances one enhances the other. A simple example would be that knowledge of the laws of physics, an intellectual enhancement, can help us create things like airplanes and cars, allowing us greater bodily movement freedom. However, this is only possible, provided someone knows how to operate these vehicles and, moreover, that there is a socially endorsed system for these operations: airports and streets, aviation codes and transportation laws, among many other things that make taking a plane or driving a car feasible. Realities that depend both on physicality and on ideas are very much intermingled in our societies. On the other hand, illness or bodily weakness can be a detriment to our proper utilization of our intellectual capabilities, like most people who have gotten the flu can confirm.

From Spinoza's quote above, we can understand that human beings feel joy when they pass into a greater degree of perfection and sadness when they decrease in perfection. In Ethics, Spinoza clarifies that joy encompasses practically all positive passions (which are analog to emotions, in our more modern understanding). In contrast, sadness does the same with negative ones and then defines various kinds of positive or negative passions (Spinoza, 1996). Further, Spinoza does not recognize any other passions except for desire itself. For us, it suffices to conceptually view joy as having positive affects and sadness as negative ones, affects being defined in this context as the experience of emotion (Nash, 2020). This concept is a direct link to modern theories of wellbeing that define the subjective dimension of wellbeing with negative and positive affects as two of its main components, like the tripartite model of subjective wellbeing (Diener, 1984), which is also incorporated as the subjective dimension of the OECD Well-being Framework (OECD, 2011b) and other objective list theories of wellbeing.

However, what is perfection for a human being? Spinoza and other philosophers after him, like Leibniz, consider that more perfect objects are so because they are more real. They act more and are less acted on, but now we ask: What is being more real for a human being?

The more perfection each thing has, the more it acts, and the less it is acted on; conversely, the more it acts, the more perfect it is.

The more a thing is perfect, the more reality it has; consequently, the more it acts, the less it is acted on. (Spinoza, 1996)

On these concepts of perfection and reality, Look points out that:

Insofar as there is the equation of reality and perfection, we might say that something is more perfect when its power is increased. Further, we can experience the positive passions derived from joy, when our power or perfection is increased; and we experience the negative passions derived from sadness, when our power or perfection is decreased […]

How do we increase or decrease our power or perfection? Here we can fall back on standard Spinozistic doctrine […]. Our power or perfection is related to the adequacy of our ideas. The clearer our ideas, the greater our power or ability to produce effects – that is, the greater our perfection. (Look, 2007)

In sum, human beings have positive affects when their ideas are clear and adequate to reality, which results in greater power, that is, a greater ability to produce effects in the world. Sadness is felt when unclear ideas that are inadequate to reality result in a diminished power or ability to produce effects in the world. As there is a unity of mind and body, clear and adequate to reality ideas allow us to interact with our environments in a way that is positive to our intentions, and not only that, but to also do it while perfecting these interactions and becoming better at the things that we do. This state of being-able-to-act while perfecting the ability to act is, in a nutshell, human wellbeing. The definition of wellbeing presented for this work can be enunciated as:

Wellbeing is the state of human beings in which their mental and bodily conditions allows them to act while gaining a greater ability to interact with their environment.

Notice that we did not include SWB in the above definition. Rather this definition encompasses both objective and subjective wellbeing because it includes the assumption that to act while gaining greater ability to interact with the environment is objectively acting in accord to the nature of human beings, their reality. Moreover, when people are in the process of gaining in this ability, they feel positive affects. When they are in the process of losing in this ability, they feel negative affects. Therefore, the subjective wellbeing component of affects of a wellbeing construct based on this definition is not only a part of an experience of wellbeing but an indicator of whether this state of wellbeing is being enhanced or decremented. This dual nature of SWB has already been pointed out by research in this area (Diener, 1984). Hence the importance of subjective wellbeing is to help understand changes in wellbeing in other components of wellbeing, even objective ones.

States, situations, and relationships under which people act that further their ability to interact with the environment should be added to our objective list construct of wellbeing. For instance, health and sufficient income certainly allow increased ability to interact with the enviroment, as do things like living safely in a jurisdiction with respect for the law, and good governance; all of which are part of the objectives list of the OECD Wellbeing Framework (OECD, 2020). However, what about roles at work, study, organizational and academic structures, ways of accountability, responsibility delegation, interpersonal relations, rules, incentives, and a myriad of other concepts found in organizations where people spend most of their waking hours than in any other place? These are more nuanced and particular to every organization. It is here where asking, “Does this organizational practice further our members' ability to interact with their environment or does it impede it?” is far more relevant to organizational change for the good, than to apply broad generic rules.

Patterns of action and subjective evaluations in organizations are key to unlocking insights about both organizational objectives and wellbeing. Furthermore, participatory organizational action is the way to effect change for the better, an approach that requires significant quantities of good information, the domain of data analytics and AI. However, we should also be mindful of the kind of object we are analyzing or intend to act upon; mainly when these objects are groups of people, hence the emphasis on a succinct definition of human nature.

II.1.3. The construct of wellbeing proposed

For the construct of wellbeing, we will use the definition of wellbeing built in the previous section as a criterion for inclusion in an objective list construct of wellbeing that will include both objective and subjective dimensions, which are formed by various wellbeing components, as can see in the diagram in **Figure 2**.

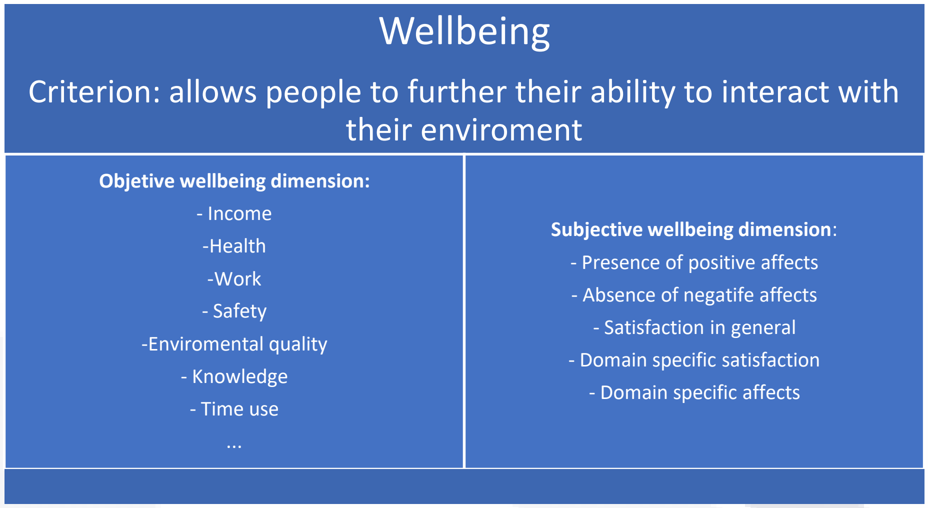


Figure 2: *Diagram of the construct of agency-based wellbeing used*

The diagram in **Figure 2** is an example of the model we will work with in the experiments recounted in this work. They are composed of three parts:

- A criterium for inclusion on the list.
- An objective dimension
- A subjective dimension

While it could be argued that this approach would necessarily lead to enumerating all the possible components of both dimensions of the wellbeing construct presented, it should be taken into account that given that the environments with which people engage have parts that are socially constructed and therefore different in different social settings (groups, organizations, regions, nations), the construct can include components that are very relevant to some individuals or organizations, and not that much to others. Given this, we argue that a wellbeing construct should be built on a case-by-case basis for a particular individual or organization and updated and modified by knowledge obtained from data collection and analysis for a new iteration or further use of the construct.

Why this on-a-case-by-case basis? The answer to this is, again, in human nature. As rational beings, part of rationality is choosing what is valuable for us according to our values and the life we intend to live. New Zealander economist G. Karacaouglu argues that an individual's wellbeing should be defined in terms of "his or her ability to live the kind of life he or she values, on a sustainable basis" (Karacaoglu et al., 2019) and emphasizes the importance of community and public policy to nurture these abilities in people. This is why this wellbeing construct can avoid the usual elitism of objective lists theories of wellbeing; it has a criterium of inclusion that requires that people exercise their two most human characteristics: their reason and their sociality. They need to find personal reasons to value a kind of life over others, and they must do it in a social process that has roots in the education their society gave them and then continue ever in processes of social negotiation, or participatory processes inside an organization or community.

II.1.4. Objective wellbeing

For the purpose of our work, objective wellbeing components are those that are relevant to allow people to act while gaining a greater ability to interact with the world and that are measured in the physical and social environment of people, as well as in their bodies, and whose values are independent of opinions and judgments of the people under study, even if the information comes from them. Objective components of wellbeing cover various domains and are constructs and objects of study in their own right; they may be socially constructed or not, or they may have some socially constructed parts and some that are not. Therefore, caution must be exercised in using them to assemble a more comprehensive construct, as in the case of wellbeing. Some examples of objective components of wellbeing are income and wealth, housing, work and job quality, environmental quality, health, knowledge and skills, safety, work-life balance, social connections, and civic engagement, to name a few of the components included in the OECD Wellbeing Framework apart from SWB (OECD, 2020).

It can be tempting to consider that what is internal (mentally or bodily) to a person and what is socially constructed is subjective; however, this is not so. SWB is subjective because it is a description of an experience and not because it is an internal bodily or mental condition or a social phenomenon. For example, the body is formed and keeps its process of homeostasis independently of how we feel about it, use our language to explain it or talk about it, and this a natural phenomenon that works independently of what we think about it (Moreno & Ruiz-Mirazo, 2015), therefore health belongs in the objective indicators. On the other hand, the mind is always formed through education, language use, participation in society (Detel, 2001), and, of course, organizations. Nonetheless, indicators such as knowledge, skills, and even mental health can be objectively measured, even if sometimes that would be difficult. As we shall see in later sections, there could be subjective assessments of every objective component of wellbeing, and these could be parts of SWB.

II.1.5. Subjective wellbeing (SWB)

According to the tripartite model of SWB popularized, broadened, and researched by American psychologist Ed Diener, which is one of the most well-known and applied, SWB is a “broad category of phenomena that includes people's emotional responses, domain satisfaction and global judgments of life satisfaction. Each specific construct needs to be understood in its own right” (Diener et al., 1999) just as is the case with the components of objective wellbeing. The components of subjective wellbeing according to this model are:

> *Positive affects: Experiences of pleasant emotions and moods.*
>
> *Negative affects: Experiences of unpleasant, distressing emotions and moods.*
>
> *Life satisfaction judgments: Cognitive evaluation or judgment about one’s life.* (Diener, 1984)

The question arises of why not consider all affects together. The affects are not considered together because there is evidence that negative and positive affects are not directly correlated (Diener, 1984). So, as positive affects diminish, negative affects do not necessarily increase, and vice versa; the conclusion is that they are separate phenomena, even if they are influenced by each other. Life satisfaction judgments can be either global, with the individuals judging their lives as a whole, or domain-specific, with the judgment being about just one aspect of life, such as work, family, health, or finances. A similar evaluation of affect balance can be done in a domain-specific way. For every component of objective wellbeing we could add at least three SWB components: satisfaction, negative affects and positive affects about that particular domain; thought, in our literature review it was found that satisfaction by individual domains evaluations are more common than affects balance evaluations.

It should be noted that SWB evaluations of particular domains are not the same as subjective evaluations of such domains; they are instead a question of how satisfactory their current state is or what affective balance that current state brings out. As an example, it is not the same to ask a person to subjectively estimate their income percentile as to ask whether they are satisfied with their income level; only the second option is within the bounds of SWB, whereas the other is a subjective evaluation of an objective component of wellbeing.

SWB resides within the individual's experience, describing how they experience the quality of their lives (Diener et al., 1999). Therefore, SWB is necessarily a self-reported measure, usually using questionaries, even as there is evidence that ML tools can predict some components of SWB, such as affects utilizing facial analysis, voice and movement patterns, perspiration, and other bodily measures (Diener & Ryan, 2008); time-use satisfaction through analyzing time use data (Marin & Ponce, 2020) and personality tests (Collins et al., 2015). Nonetheless, the predicted value remains an indicator of an experience, which is always subjective. Machine learning tools can merely learn how to recognize the presence of these subjective states by employing datasets categorized by people who have experienced them.

II.1.5.1. Affects

Affects, both positive and negative, are essential components of SWB. Affect is “a term used in psychology to denote the broad field of emotional and mood-based experience of the human subject” (Nash, 2020). Sometimes the term affect is used as a practical synonym for emotion, but it is more correct to say that an affect is the experience of an emotion. Some authors use affect to denote emotions or feelings that can be observed in a person's behavior, such as facial expressions or body language (Leeth, 2011); that can be conterminously apparent with the emotional experience. Therefore, we say that we are undergoing an affect, or are *affected* when we experience an emotion; and we show it in our behavior due to involuntary or voluntary responses.

As we have already seen, affects are the “online evaluation of the events that occur in [people’s] lives” (Diener et al., 1999). Positive affects are experienced when people undergo their ability to interact with their enviroment enhanced, and negative affects when it is decremented (Look, 2007). Therefore, affects are an immediate evaluation of what is currently happening in a person's life. A balance between negative and positive affects is essential for a state of SWB. While there is no consensus on the definition of emotions, modern sciences tell us that emotions can be characterized as “physiological and behavioral responses to an important environmental object or event, typically mediated by the autonomic nervous system and subcortical brain structures. These responses are combined with cortically-mediated modulation and interpretation” (Waraczynski, 2009).

Nevertheless, what counts as "important environmental objects"? Who decides that and on what basis? We can return to the definition of affects as experiences (of emotions) that result from acting, experiencing, or anticipating an enhanced or diminished ability to interact with the world (Look, 2007). Given that the criteria to distinguish between things that enhance, diminish, or have no effect on personal wellbeing are necessarily of a rational nature, for they can discriminate between different kinds of experiences and consequences, we can consider emotions as thoughts accompanied by a limbic response from the autonomic nervous system sounding the alert to a perceived danger or opportunity through an evaluative, and thus rational, judgment and not merely experiences of "non-reasonable movements, unthinking energies that simply push the person around, without being hooked to the ways in which she perceives or thinks about the world" (Nussbaum, 2001). Therefore, emotions can be rationally engaged with.

This clarification on the nature of emotions is necessary because, for SWB research, there may be a need to work with emotion as alerts to be taken into account that something important is happening in an organizational setting, mainly if these emotions arise while performing actions in this setting; which would heighten the probability that the organizational environment has prompted the emotional state and it is the purview of management to address it. It should be noted that this is not the same as working with "feelings," which are the interpretation of emotions (Waraczynski, 2009) and are outside the scope of this work and its objectives. We only need to consider the presence or absence of experienced emotions, or affects, because their presence tells us something important about wellbeing being enhanced or diminished in a particular environment or situation. What interpretation a person makes of his emotional experience is wholly in another domain than those considered in this work.

Suppose a negative affect arises while performing an action. In that case, this action likely includes factors beyond the control or beyond the person's strengths that are, nonetheless, important to the person acting or have significant consequences. This situation can be compounded if the action is something that does not lend itself to be a learning experience or one of perfecting one's practice in the activity and other related ones: something tedious, repetitive, and unconnected to real needs and personal strengths. Suppose a positive affect arises while performing an action. In that case, it is likely that this action is being done under control or has been planned by or with the participation of the person performing it, and it is something that lends itself to be a learning experience or to perfecting the practice in it or other related activities. Therefore, in an organizational setting, affects not only become indicators of members being in the process of enhancing or decreasing their wellbeing, but also of members being in states conducive to high performance, such as flow, or intrinsic motivation, which is “the motivational component of people's natural proactivity, [referring] to doing an activity because it is interesting, enjoyable, and inherently satisfying of their basic (i.e., innate and universal) psychological needs” (Moller & Deci, 2014). Again, affects are not unconnected to an objective reality relevant to the organization. Affects, not feelings, are the key concept for wellbeing change monitoring, though it should always be kept in mind that actions are how such changes are brought about.

However, not all affects are about the action being performed at the moment. What about emotions that come because of mere thought? Apart from pointing out that there could be a triggering influence from the social and organizational environment in people's thoughts, we should clarify that to think is an action too, and one we do both with our minds and bodies, as Spinoza would argue, and as modern science and study of the nervous system tell us. A thought that elicits a negative emotion comes because we are thinking about something outside our control that could have a substantial negative impact on our wellbeing, and similarly with positive emotions, except that it means that the impact on our wellbeing would be positive. Nonetheless, if these phenomena became relevant to performance, this situation is still of interest to the organization and its members, as professional psychological interventions can substantially decrease intrusive negative thoughts. The availability of these can be valuable parts of both a wellbeing and wellness program.

The importance of adequate affective balance is also important because there is evidence that positive and negative emotions are related to job performance (Aeon & Aguinis, 2017), and maintenance or decline of health (Trudel-Fitzgerald et al., 2017); and that greater positive emotions are associated with cardiovascular health (Boehm et al., 2020); and although causality is not clear even in Boehm et al’s longitudinal study, an action-based approach to emotional wellbeing would be focused not in dealing with the emotions themselves (which would make it necessary to deal with their interpretation) but the issues that caused them. These links between emotions, health, and job performance let us see how the cycle of body and mind closes being a single reality, a human person that can grow towards perfection by acting, by doing with the body what clear thoughts adequate to reality lead to and using emotional experiences as guidance to finetune our thoughts and actions according to reason.

II.1.5.2. Satisfaction

The final part of the subjective wellbeing model is satisfaction. This component "is a cognitive judgment evaluation of one's life. As such, it may be indirectly influenced by affect but is not itself a direct measure of emotion" (Diener, 1984) as it is based in more explicit judgements than affects. Satisfaction is a more consistent and stable variable than affects (Diener & Ryan, 2008). Also, satisfaction evaluations can be made on particular areas of life, such as work, love life, friendships, finances, health, and others. Some of these particular satisfaction evaluations correlate more with general life satisfaction, such as love life and friendships. General life satisfaction is more correlated with personality types as assessed by specialized tests than positive affects, but not more than negative affects (Emmons & Diener, 1985).

Given the relative stability of general life satisfaction and its correspondent satisfaction judgments in specific areas such as work satisfaction, the interest of these measures in organizational settings is more long-term than the affects components. They can serve less to evaluate situations with immediacy and more to tailor the organizational systems to its members' needs and personalities to increase life satisfaction over time. Whether satisfaction is correlated with having high levels of ability to interact with the enviroment is in question in this work.

II.1.6. Wellbeing in organizations

Let us restate the definition of wellbeing that we will be using in the rest of this work:

> **Wellbeing is the state of human beings in which their mental and bodily state allows them to act while gaining a greater ability to interact with their environment.**

The constituent parts of this definition of wellbeing can be explained as follows:

- **Wellbeing is** a state that human beings can be in, and this state has two interdependent components:
    - **The state of the body:** It allows us to move, accomplish an action, and interact with our environment and other people. However, it also allows us to think clearly and participate in collective and individual endeavors.
    - **The state of the mind:** That the ideas, concepts, and mental processes are clear and adequate to reality.

- **To act while gaining greater ability to interact with the enviroment:** To act intentionally in the world adapting it to us, and not just adapt to it, thus having the capacity to change the environment.

It would be a mistake to think that we can individually attain increased wellbeing. This fact is because part of a state of mind adequate to reality includes acknowledging the social construction of things like desires, needs beyond the bodily basics, recognition, social institutions, and even the practices we use to take care of our bodies and minds, as well as every action demanded of every role that some social group or organization recognizes.

Social construction “refers to the processes that accompany the joint acts through which the social world is created” (Grills, 2017). Briefly, socially constructed objects would currently only be found in a world with human beings; the rules they operate by are decided socially and can only be changed socially. These socially created rules still need to comply with things like logic and pay attention to the laws of nature when engaged with the natural world. However, they are more flexible than, say, the law of physics, which can only be learned and considered in our endeavors instead of designed or constructed. Socially constructed things include money, law, ethics, governments, organizations, arts, and other realities.

Organizations and the rules and roles under which they operate are socially constructed, as there are no natural laws independent of humans that govern how organizations should operate. Many realities that human beings experience are socially constructed. Most human experiences and processes have socially constructed components and parts ruled by natural processes independent of what human beings might think about them. For example, the need for food or physical nourishment is not socially constructed as it is part of what every known living being is subject to; but what, when, how and with whom to eat are socially constructed parts of the experience of nourishing our bodies with food.

It is essential to understand this because the ways organizations work are socially constructed and, therefore, subject to change to better fulfill some objectives, including wellbeing. It does not mean that organizations can be changed on a whim or are subject to fictitious and unnecessary rules. Rules enacted, recognized, and followed by a group of people are very real, and they can even be necessary for the organization to continue its function and realize its objectives. Socially constructed things are real, they exist, just not in the same way that things like the laws of physics or a star (that behave the way they do independently of what human beings think of them), as opposed to things like laws or rules, that can be changed or even eliminated, but apply with full force across their jurisdiction, and originate different consequences if respected or ignored. Nonetheless, socially constructed things and those that are not can make their causal influence felt on organizations and their members.

How things exist is the area of study of Ontology and beyond the scope of this work, but suffice it to say that when dealing with constructs with a high level of social construction, such as wellbeing and operating in environments that have a high component of social construction such as organizations it is crucial to keep in mind two ideas when considering data about objects and people: the distinction between objects and processes that are socially constructed and those that are not, and that some processes and objects can have some components that are socially constructed and some that are not. This distinction will help to spot opportunities that are usually missed during an analytical process by seeing new possibilities with situations that seemed unchangeable before and problems that seem intractable but are not, and to avoid confusion when processing the data produced in organizations.

Organizational change is hard-won and includes great quantities of thought, research, and experimentation, along with the application of extensive knowledge about how people behave in an organization. It is hard work. Furthermore, that is why powerful sets of tools like AI and data analytics become relevant at this juncture. Particularly if the object to enhance is the wellbeing of human beings that are a blend of socially constructed objects and processes and those subject to the uncompromising natural laws discovered and explained by sciences such as physics, biology, physiology, and medicine.

II.1.7. Wellbeing as data analytics an AI target

Every piece of wellbeing we have in our lives, from having food on our tables and the shelter of our houses, to the recognizement of our efforts and an a plethora of other things that can be necessary to our happiness, and that can be summarized in an objective list theory of wellbeing, could not exist without some organized group of people engaging the world to adapt it to our needs, be it a state, a school, a company, markets, groups, nations, our families and pretty much every organization we belong to. Moreover, while engaging the world, these groups necessarily use symbolic language (one of the things that are natural to humans and only to humans) and define concepts and relationships between these concepts and the roles individuals can play in order to become an organized whole (another of the things that are natural to humans) that can confront the natural and social world and adapt it to their needs. Therefore, not considering this sociality of the ideas that inhabit and form our minds would be inadequate: it would amount to holding ideas that do not conform to reality. It would also lead to being unable to change the practices of organizations that impede their stated objectives, be it profit, the common good, or wellbeing.

Therefore, it is imperative to recognize wellbeing, and the lack of it, as a social phenomenon, both because to do so is adequate to reality and because, if so, the study of wellbeing lends itself to the approaches of data analytics and AI, both disciplines that are fit for the exploration of wellbeing in organizations because the insights to solve the problems in this area are most likely locked away in the massive datasets that already exist in many organizations, and in the perceptions of their members about their wellbeing which need to be taken in the social context of the organization. Organizations are full of people engaged in action, which is the context in which enhancement of wellbeing or its diminishment is experienced; and modern organizations create detailed trails of data for most of these actions that form massive datasets. The question is… are these actions creating wellbeing, allowing us to better interact with the world on human terms? This a question that can be answered in both a descriptive and prescriptive way by AI tools.

II.1.8. Connecting the wellbeing construct with other compatible definitions of wellbeing

As we have already mentioned, wellbeing has many definitions these days. Now that we have the definition by which this work will judge all others, we can see some of these relevant to this study. In recent decades, there have been significant advances in the study of wellbeing and calling attention to the importance of emphasizing wellbeing in our organizations and societies. Many researchers and authors from different areas of science and philosophy have developed theories of wellbeing, and even governments and international organizations have come to develop their constructs to define wellbeing.

This state of affairs means that there is a richness of studies and data available that operationalizes definitions of wellbeing that differ from those explained in the last section of this work. However, this need not be an insurmountable obstacle to take advantage of this information provided that we consider the differences and similarities between different definitions of wellbeing to know what further preprocessing and processing needs to be done to the data to extract from it further insights. Also, the experiences and resources developed by researchers in this area have provided ample information on designing the experiment presented in other chapters of this work.

For this dissertation, four important theories of wellbeing are deemed compatible with the one presented here, and not only that, but they are also immediately relevant as these are ones with which we have engaged in research because of their ample use or richness of the data created by past research and data collection taking these definitions or constructs into account. These theories are listed and explained below:

- **Tripartite theory of subjective wellbeing** is one of the most used and well-researched theories. Diener's model is the modern heir to the tradition of the importance of positive and negative effects as seen with philosophers like Spinoza and Hegel. It defines the subjective dimension of wellbeing with three components: positive affects, negative affects, and life satisfaction (Diener, 2000). Articulated in a seminal article in 1984, it is one of the first systematizations of the idea that what people think, and feel is essential to evaluate their wellbeing. This theory fits seamlessly with the definition posited in this work and its three components are included in the objective list construct presented in this work as seen in **Figure 2**.
- **OECD Better Life Index:** Created in 2011, this index includes eleven dimensions of wellbeing, both objective and subjective: housing, income, jobs, community, education, environment, governance, health, life satisfaction, safety, and work-life balance. This model incorporates various economic and social metrics and the life satisfaction component of subjective wellbeing as developed in Diener's tripartite theory of subjective wellbeing (OECD, 2019). This index was developed in response to the Stiglitz report that concluded that standard macroeconomic statistics like gross domestic product (GDP) failed to give an accurate account of people's current and future wellbeing (OECD, 2011a; Stiglitz et al., 2009). All the dimensions, or components included in the index can also be included in the objective list of the construct presented in this work.
- **Intergenerational wellbeing:** Defined as the objective of sound public policy, wellbeing is defined as “an individual’s ability to live the kind of life he or she values, on a sustainable basis” and it is posited that “good public policy contributes to individuals’ wellbeing by enhancing their capabilities and opportunities to do so” (Karacaoglu et al., 2019). This

definition is very similar to the one presented in this work, as it equates wellbeing with the "ability to live the kind of life one has reason to value" (Karacaoglu et al., 2019) and it is in fact the connecting tissue between the wellbeing construct presented in this work and the OECD Better Life Index via its emphasis on reason. The intergenerational wellbeing definition differs in the way it does not say that this ability must increment as it is said in the definition of wellbeing in this work, only that bringing about the possibility of its increment is the objective of good public policy. However, as this definition is constructed from a public policy and economics viewpoint and the one presented in this work from organizational and data analytics, they are more complementary than different.

- **Wellbeing Five:** A multidimensional construct of wellbeing that is composed of five statistical factors gleaned from surveys and studies done by the Gallup Foundation. According to this theory, a life of wellbeing consists of a balance of taking action and care of five dimensions of life: work or career, finances, physical, community, and social. A summary of the recommendations of the model is shown in **Table 1**.

Table 1: *The statistical factors of Wellbeing, according to the Wellbeing Five model and three recommendations or call to actions for each one (Rath & Harter, 2010)*

Work/Career wellbeing 1. Every day, use your strengths. 2. Identify someone with a shared mission who encourages your growth. Spend more time with this person. 3. Opt into more social time with the people and teams you enjoy being around at work.	**Financial wellbeing** 1. Buy experiences, such as vacations and outings with friends, or loved ones. 2. Spend on others instead of solely on material possessions. 3. Establish default systems (automated payments and savings) that lessen daily worry about money.
Physical wellbeing 1. Get at least 20 minutes of physical activity each day, ideally in the morning to improve your mood thorough the day.	**Community wellbeing** 1. Identify how you can contribute to your community based on your personal mission.

2. Sleep well enough to feel well-rested (generally seven to eight hours) but not too long (more than nine hours). 3. Set positive defaults when you shop for groceries. Load up natural foods that are red, green, and blue.	2. Tell people about your passions and interests so they can connect you with relevant groups and causes. 3. Opt into a community group or event. Even if you start small, start now.
Social wellbeing: 1. Spend six hours a day socializing with friends, family, and colleagues (this time includes work, home, phone, e-mail, and other communications.) 2. Strengthen the mutual connections in your network. 3. Mix social time with physical activity. For example, take a long walk with a friend so you can motivate each other to be healthy.	

II.2. Time use

Time can be seen as a resource that, "although equitably distributed, is a fundamental source of expression and social differentiation. As time has been used in the past, it accumulates to embody human capital" (Gershuny, 2015) that includes social, cultural, economic, and even health capital. In other words, the use of time is directly connected to wellbeing. Much of the current situation and potential of individuals and societies is due to how time has been used, having made "investments" of it in various activities in accordance with personal and social values and intentions. Hence the importance of the study of the use of time, to the point that it is an important component of the study of the wealth or poverty of societies and individuals as an inclusive concept (United Nations Development Programme, 2019), including dimensions of personal wellbeing.

II.2.1. Definitions relative to time-use

Time is "a fundamentally scarce, finite, and irreplaceable resource available to all men, women, and children in equal amounts of 168 hours per week throughout their lives whose use largely determines the progress, achievements, and wellbeing of individuals, families, communities and societies" (Ironmonger, 2018). Economics being the science of the management of scarce resources, we could talk about an economy of time, which can be used in different ways. However, unlike money or other material resources, time is an intangible reality that is unmanageable itself. As we will see in the following sections, when we talk about time management, we are talking about managing the use of time; that is, we do not manage time: we manage activities and assign them to periods in which they can be done. However, time use should not be confused with time management, time use is merely the assignation of time to alternative uses, such as work, sleep, leisure, and others (Ironmonger, 2018).

Time-use studies show how people assign their time to different activities and, at minimum, show the activities in which people use their time on a weekly or daily basis; maximally they show what they are doing, with whom, where, and how they feel minute to minute (A. S. Harvey & Pentland, 2002). Most time-use studies depend on different kinds of time-use surveys that ask the respondents about how much of their time they assign to different activities through some predefined representative period. However, in data-rich environments such as organizations that use software to manage activities or do monitoring, the data required for time-use studies can be obtained simultaneously with the activities being done in an automated way.

The electronic monitoring of activities need not be a full-blown surveillance system. These wide-ranging systems are developed and implemented with the explicit purpose of surveilling users, employees, students, or members of an organization to know what they are doing within the period in which they are working under or for the organization with a predefined level of detail. These systems can include cameras, facial and emotional recognition systems, surveillance software such as keyloggers and mouse trackers, and even sensors to measure values related to the state of the body. It should go without saying that such exhaustive monitoring raises privacy and even wellbeing concerns, and that is why one of the purposes of the research presented in this dissertation is to work with data whose collection is minimally intrusive as with traditional time-use surveys.

II.2.2. Time-use surveys

Time-use surveys are the systematic application of measurement instruments to collect data about how a group of people uses their time. This is done by measuring “the quantity of time people dedicates to various activities” (Eurostat, 2019). The measurement instruments for time-use surveys can have different designs. Some come in the form of a diary, in which people input their activities while going about their day or at the end of it. However, the most common measurement instrument for time-use surveys is the stylized questionnaire that includes a predefined set of questions asking about particular activities (Committee on National Statistics National Research

Council, 2000), and the respondents answer how much time they assigned to each activity in a representative period, such as last week, last month or other convenient periods.

II.2.3. Satisfaction in time-use

One very important measurement of SWB, directly related to the level of agency, or capacity to act intentionally (Schlosser, 2019), that people have, is how satisfied people are with their patterns of time use. A question such as "How satisfied are you with your academic activities or your work?" is different in a very important way from "How satisfied are you with the quantity of time allocated to your academic activities or work?". In one question, we are asking the respondent to evaluate the activities themselves according to their intentions and expectations, hence asking about satisfaction, while in the other, we are asking to evaluate if the right amount of time is being allocated to such activities irrespective of how these activities could be evaluated.

For example, some people might not even be able to answer a question such as "How satisfied are you with your academic activities" if they do not have any. But they may be able to answer that they are satisfied with not allocating time to academic activities if their intentions do not include these activities at the moment, or they might answer that they are, in fact, unsatisfied with not having academic activities. Dissatisfaction in time use indicates a problem with agency: a person wants to engage in certain kind of activities for a given period but has been, for one or other reason, unable to. The possible combinations of satisfaction in time and in the activities done is shown in **Table 2**.

Table 2: *Possible combinations of satisfaction with activities and satisfaction in the time assigned to them*

	Satisfaction with the activity	**Dissatisfaction with the time use**
Satisfaction with the time use	Considers the activity meets their expectation and is in line with their intentions and can dedicate to the activity an amount of time that is adequate.	Considers the activity meets their expectation and is in line with their intentions; however, they are unable to dedicate to the activity an amount of time that is adequate. They may want to assign less or more time to the activity.
Dissatisfaction with the activity	Considers the activity does not meet their expectation or is not in line with their intentions, but they can	Considers the activity does not meet their expectation or is not in line with their intentions, and they are unable

	dedicate to the activity an amount of time that is adequate.	to dedicate to the activity an amount of time that is adequate. They may want to assign less or more time to the activity.

Also, agency is not the only variable that is linked to time-use satisfaction, there is empirical evidence that the subjective experience of time-use influences decision-making, behavior and other subjective states of people, particularly those related to anxiety over time (Boniwell, 2006). People's expectations and evaluations of time use are then linked in an interdependent way to SWB.

II.3. Time management

Time management is of great interest to the development of intelligent tools to preserve or enhance wellbeing in organizational settings because, as we have seen in the previous section, it is only through actions born of ideas adequate to reality that both organizational and wellbeing objectives can be met, and time management practices are a good opportunity to clarify ideas and values as well as to refine paths of action. Also, the need to collect data in which to train accurate AI tools demands that thought be applied to the kind of data necessary, and therefore to the goals and value expected.

Different courses of action, and even sets of values, can compete for different scarce resources, but one of the scarcest and always needed is time. As already mentioned, and Claessens argues, the term time management itself is a misnomer, as "time cannot be managed, because it is an inaccessible factor. Only the way a person deals with time can be influenced" (Claessens et al., 2007), which leave us one more time in the realm of actions or behaviors, which as we have seen can be measured objectively to produce data amenable to an analytics approach which can result in the kind of intelligent tools developed for this work. It is important to note that what to measure and how to interpret the data depends on the socially constructed environment and intentions of the organization in which the practice of time management is engaged in.

II.3.1. Antecedents of time management

We have already mentioned that time is can be considered a "scarce resource" and that, because of this, we can work with an economy of time, which is a resource "whose use largely determines the progress, achievement, and wellbeing of individuals, families, communities and societies" (Ironmonger, 2018). And not only that, the way people use time accumulates to form capitals that are not only economic or financial, but cultural, social, and other embodied kinds of capital, as health could be characterized (Gershuny, 2015). Given this important role in human affairs, it could be expected that the management of such an important resource to be an influential discipline with a great corpus of literature and research and great innovations powered by the tools of the digital age. However, this is hardly the case.

The notoriety of time management research and applications greatly increased during the second half of the 20th century. Most of the time management literature was about its benefits and implementation and very little actual research was published to back those benefits up (Macan, 1994). Nonetheless, time management was incorporated in many kinds of tools and management philosophies used to increase efficiency during this period. However, as Aeon and Aguinis (2017) point out, time management as a systematization is not a novel approach born out of the advances of industrialization, and actually predates the Gregorian calendar and the mechanical clock with the rule of St. Benedict from the V century being one of the oldest examples still used in some communities (Aeon & Aguinis, 2017). And while information is less reliable the further back in time the role of time management is considered, it is at least not implausible that great architectural projects such as the pyramids in Egypt and Mexico, and the great temples of antiquity, to name a couple of human endeavors that we can still appreciate thousands of years after their completion, required a system of time management for the labor force needed to build them.

As part of our own literature review, we realized that one of the first things a researcher delving into the time management literature can notice is that while there is rich ecosystem of time management systems ready to be used, there is hardly a time management studies cohesive discipline. This a very interesting fact given the importance of time as a factor in practically all human endeavors. Rather, time management studies and applications are scattered throughout different disciplines that use different interpretations of time use and management (Aeon & Aguinis, 2017). This situation is compounded as each discipline approaches the issue of time from a different perspective and with different intentions. For the purposes of this work, this is not a disadvantage, given that we argue that knowledge is best constructed by contrasting different takes with different intentions on the same concept we want to have knowledge about; which is the way we can construct a data analytics application fed from different sources of data produced by tools with different intentions in mind.

As relatively little time management research has been conducted even during the XX and XXI centuries, we still do not have enough data to do metanalytic work (Claessens et al., 2007), but enough to have a relatively broad perspective of the impact of time management applications in many factors studied by the disciplines in which these time management techniques were examined, including organizational issues and wellbeing. It was not until 2017 that Aeon and Aguinis published another review of the existing literature and integrated their insights using the concepts of time structures, time norms and the role individual differences play in time management (Aeon & Aguinis, 2017) in order to make sense of some of the contradictory findings of the previous decades.

Aeon and Aguinis (2017) mention several disciplines with a considerable corpus of time management literature and applications such as Sociology, Psychology, Education, Management, and Consumer Behavior and some subdomains in which the insights of these disciplines are applied, such as work-life conflict, job performance, cross-cultural management, stress, creativity, and life satisfaction. While the approaches across these disciplines and domains differ according to their intentions, there is also two critical outcomes sought across all: performance and wellbeing (Aeon & Aguinis, 2017). These are precisely the two variables that we are the most interested in.

It is in this moment that we arrive to an important fork in the road of time management as practiced nowadays: performance or wellbeing. This is not to say that we cannot design a time management method or tool that has both concepts at the core of their design, but this has been very rarely the case. In fact, most time management approaches have dedicated their efforts to enhance performance and efficiency and have let wellbeing enhancement be a byproduct of that process. Even what sometimes passes as a wellbeing approach is mostly a getting-things-done approach whose goal is to either have more free time, producing peace of mind, or to squish wellbeing promoting activities such as socializing, exercise or cooking healthy meals into an already packed day plan.

This is not surprising given that most time management systems were developed for an industrial or commercial setting in which getting-things-done in time and effectively is crucial, and this approach has seeped into different settings like education and wellness. But this is also the reason time management has somewhat of a reputation as an oppressive tool; and a free or blank calendar or day plan has become a symbol of freedom and happiness, thought this is hardly the case.

II.3.2. Definitions of time management

As it happened with the concept of wellbeing, and even those related to data analytics such as big data and data mining, there are many definitions of time management. As we have seen, this is due to the scattering of time management research across many domains and disciplines. Both Claessens et al. (2007) and Aeon and Aguinis (2017) gathered some relevant definitions.

- Time management…
    - Involves the process of determining needs, setting goals to achieve these needs, prioritizing and planning tasks required to achieve these goals.
    - Are techniques for managing time.
    - Is a technique for effective time use, especially having enough time to accomplish the many tasks required.
    - Is planning and allocating time.
    - Is the degree to which individuals perceive their use of time to be structured and purposive.
    - Is a way of getting insight into time use.
    - Is a technique to increase the time available to pursue activities.
    - Are practices intended to maximize intellectual productivity.
    - Is an application of self-regulation processes in the temporal domain.
    - Is coping behavior in at-risk populations.
    - Are self-regulation strategies aimed at discussing plans, and their efficiency.
    - Are the use of procedures that are designed to help the individual to achieve his or her desired goals.
    - Are ways to assess the relative importance of activities through the development of a prioritization plan.
    - Are clusters of behavior that are deemed to facilitate productivity and alleviate stress. (Claessens et al., 2007)

- Time management…

- Is a combination of time assessment, goal setting, planning, and monitoring activities.
- Are self-controlled attempt to use time in a subjectively efficient way to achieve outcomes. (Aeon & Aguinis, 2017)

Claessens et al., as a conclusion of their review present their own definition of time management, considering time management as behaviors.

> **"Time management is defined as behaviors that aim at achieving an effective use of time while performing certain goal-directed activities"** (Claessens et al., 2007)

As we can see, this definition again puts forward the importance of actions, or behaviors, and clarifies that these behaviors are aimed at an effective use of time with some intention in mind: meeting goals or objectives. And so, we see that in the context of time management the "use of time is not an aim in itself and cannot be pursued in isolation. The focus is on some goal-directed activity, such as performing a work task or an academic duty, which is carried out in a way that implies an effective use of time" (Claessens et al., 2007) and all this only makes sense in the socially constructed contexts of organizations and societies. This mirrors the analytics process used to build AI and data analytics tools; it is not done for its own sake, but with some clear goal in mind relative to a value that is intended to be generated, and the decision to name or choose that value is always a social one.

Also, the authors classify the behaviors mentioned as:

> ***Time assessment behaviors,*** *which aim at awareness of here and now or past, present, and future […] and self-awareness of one's time use […], which help to accept tasks and responsibilities that fit within the limit of one's capabilities.*
>
> ***Planning behaviors,*** *such as setting goals, planning tasks, prioritizing, making to-do lists, grouping tasks […] which aim at an effective use of time.*
>
> ***Monitoring behaviors,*** *which aim at observing one's use of time while performing activities, generating a feedback loop that allows a limit to the influence of interruptions by others […].* (Claessens et al., 2007)

On the other hand, Aeon and Aguinis (2017) argue that given the different intentions of different research areas, the variety of definitions of time management is to be expected and it is for this reason that “we need a definition that subsumes, integrates, and applies to a wide range of disciplines […] a person-centered perspective in which we conceptualize individuals as proactive and intentional agents. In line with this perspective, we contend that individuals make decisions about how they allocate time” (Aeon & Aguinis, 2017). It is to be noted that Aeon and Aguinis already conceptualized individuals as proactive and intentional, that is, with agency, at least potentially. Therefore, they present a definition of time management as:

> ***[Time management is] a form of decision making used by individuals to structure, protect, and adapt their time to changing conditions.*** (Aeon & Aguinis, 2017)

Aeon and Aguinis (2017) characterize this definition as follows:

> *This definition is consistent with an agentic perspective of time* (Granqvist & Gustafsson, 2016). *Indeed, calendars, schedules, holidays, semesters, clock time, and weekends are not “brute physical facts” […]; rather, they are social constructions subject to change and negotiation […]. At the individual level of analysis, people are arguably free to organize their time as they see fit […] by drawing on existing time models […] or creating their own unique time structures […].* (Aeon & Aguinis, 2017)

The only quibble we have with this definition, is its emphasis in individual decision making. Given that the authors go to great lengths to make this definition an agentic perspective of time, or in other words, to consider that the individual has agency to “structure, protect and adapt their time” (Aeon & Aguinis, 2017) we can use it while adding that the only way such agency can be had by the individual is in the context of participating in the construction of time norms and time structures of a social group the individual is a member of, which can be an organization, which to the authors’ credit is expressed in the discussion of the results analyzed in their review. This is because it is the organization that sets objectives, roles, rules, expectations, responsibilities, limits, different levels of importance of different tasks, and even gives relevance (or not) to concepts such as punctuality and delivery on time.

It is also the organizational culture that resolves conflicts between different values such as between the quality of deliverables and deadlines, or time spent in the office and work done. It is only against this organizational background that the individual can manage their time, and agency is obtained through participation in the construction of the organizational culture, and in decisions about the values and objectives of the organization which, as we have seen, it's a facet of having greater ability to engage, or wellbeing as defined in the previous section of this work. In a time management system that has wellbeing as a core value, setting time structures and time norms can hardly be the exclusive domain of managers, because that would hinder wellbeing for all the members of the organization, including managers.

Time structures and time norms are part of a set of four research perspectives that Aeon and Aguinis (2017) define to better understand the dynamics of time management in organizations:

> ***Time structures*** *are "those external aspects . . . that can be described more or less reliably by an independent observer" […], such as the timing, frequency, sequence, and duration of events […]. Business hours, project timelines, cleaning schedules, and holidays exemplify time structures; they are explicit and formalized. Time structures affect individual time management by laying out a system around which people can organize their time.*
>
> ***Time norms****, [are] intangible and shared patterns of expected temporal activity […] that become more salient once they are breached. They constrain time management behavior through social pressures. […] What distinguishes time norms from time structures is the moral connotations attached to time norms […].*
>
> ***Individual differences*** *in personality, values, and beliefs are known to influence various organizational outcomes such as job satisfaction and performance […]. Similarly, individual differences—specifically, individual differences in time attitudes, beliefs, and preferences […]—play a critical role in affecting time management outcomes.*
>
> ***Temporal decision making****—the examination of how people make time related decisions—enhances our understanding of time management. The time management literature implicitly adopts a "rational time manager" model, according to which individuals make optimal time decisions. Other fields, such as economics and strategic management, have also traditionally treated individuals as rational, optimal decision makers. However, developments in behavioral economics […] and strategy-as-practice […]have upturned some of these assumptions.* (Aeon & Aguinis, 2017)

Perlow et al. provides an example in which time management tools were introduced to an organization, but the engineers who made the most of these saw a drop in the evaluations from their supervisors even as they completed all their work in time instead of staying late to do it. It so happened that in this organization long hours and staying late in the office were interpreted as signs of commitment and being hard working (Perlow et al., 2016). This example is also reiterated in the Aeon and Aguinis review (2017).

In the example mentioned above we see that the agency of the engineers was limited by their organization. They could certainly individually organize their workday to be more time efficient with intentions such as being home in time, and leaving all their tasks complete, but the organization did not recognize this as a good thing. Instead, their standing with their supervisors suffered and all their efforts could result in negative consequences from further negative evaluations, deteriorating standing with their peers from “not doing as we all do”, and even loss of promotions or other problems leading to termination of their employment. They did not actually have the sufficient agency to better manage their time without first addressing the time norms and time structures embedded in the culture of their organization. This is something that needs to be done with participation of all the stakeholders and cannot be done by fiat even by managers, as even good managerial practice in the case of changes is to get the whole team onboard with them.

In the example of the engineers, we can see that if the study didn’t include the perspectives of time norms and time structures first, and only later considered individual differences and the role of the temporal decision making, the research would generate a result in which time management is negatively correlated with good work performance (as evaluated by supervisors). Once the time structures of the organizations are taken into account, such as deadlines, business hours and schedules; and time norms are made explicit, then the issue of work during business hours being a bare minimum expected and long hours being the ideal in order for supervisors and peers to expect deadlines being met is revealed as a factor that mediates the work performance evaluation and, it should be taken into account to either qualify the results of the research as only relevant to organizations with this kind of time structures and norms, or the study should include a process of change of these norms and structures. Intelligent tools for time management should be designed with these time norms and structures in mind or be capable of learning these time structures and norms in order to at least preserve wellbeing and performance, and ideally, should be capable of recommending change in these structures and norms in order to enhance wellbeing and performance.

Individual differences, on the other hand, become relevant to study why some people under the same time structures and norms make better or worse use of time management training and tools; and temporal decision making can help to model the decisions of individuals and groups of people according to how this process occurs. Both perspectives of time management research can be greatly benefited from a data analytics approach, thought these are not the focus of this dissertation.

As the definitions of Claessens et al. (2007) does clarify, time management is only relevant while performing goal-directed activities. It can be the case that the individual has limited agency to choose his activities while not participating in the social construction and negotiation of an organization rules, roles, and goals, but once these activities or responsibilities are negotiated or at least made explicit in the context of the organizational culture, individuals may have greater agency to manage their time. Applications of data analytics, including intelligent tools of time management, can help with these negotiations and clarifications, and with the construction of an agentic time management system in which all stakeholders can give each other greater agency and, therefore, greater possibilities of wellbeing; along with greater possibilities of carrying to completion the objectives of their role in the organization.

Even in the case of an individual managing their own free time, Aeon and Aguinis (2017) qualify such freedom with “arguably” because, indeed, it can be argued that the very concept of managing time is a social construct, and the individuals can only choose between the options of time management that their education and culture has made known to them, to manage their time to accomplish objectives expected of them according to their role in society. Aeon and Aguinis (2017) present different levels of analysis of time norms and structures at team, organization and country level, these are presented in **Table 3**.

Table 3: *Summary of why time structures and time norms affect time management at different levels of analysis such as team, organization, and country (Aeon & Aguinis, 2017)*

Level of analysis	Time structures	Time norms
Team	Teams agree on rules via consensus. Such rules can be time-related (e.g., work starts at 8 a.m. sharp) and influence the time management behaviors of individual members.	Teams develop implicit time-related norms that can constrain individual time management (e.g., time is precious in our team, and wasting it will be severely punished).
Organization	Organizations use time structures (e.g., business hours, project timelines) to standardize and control individual time management practices.	Through socialization and reward systems, organizations instill time norms in employees to channel their individual time management practices toward organizational goals.
Country	Cultures and institutions have different ways of organizing time—hence the differences in time zones, business days, and other time structures across countries and institutions. Individuals such as travel executives and global entrepreneurs must be mindful of those differences to seamlessly coordinate their global operations.	Different cultures and institutions have wildly different norms with regard to time. To avoid conflict, the frequently traveling employee must be time-culturally savvy.

We will keep in mind both definitions, the one put forward by Claessens et al. (2007) and Aeon and Aguinis' (2017) one. The emphasis on time management being some kind behaviors of the Claessens et al. definition will help us not to lose sight of what actions and the data generated by them correspond to time management activities. On the other hand, the emphasis on decision making and agency of the Aeon and Aguinis definition will help us link these behaviors with the definition of wellbeing posited in the last section: the more agency to make decisions the members of an organization give each other to manage their time, the more ability to interact with their enviroment is possible for this group of people. Wellbeing is not the same, nor does it follow from, the ability to manage one's own time, but this ability is clearly a prerequisite for enhanced wellbeing.

II.3.3. Intelligent time management

No organization lacks some system of time management. Even those without formal time management systems manage time on the fly using the methods and resources that members of the organizations find or test during the first times the need to manage time arises. These methods used come from the culture, education, and experience of the members of the organizations; they develop organically but ossify quickly into a system of minimum effort tolerated. In other words, when an organization lacks a formalized time management system, they do what the research team now calls time management by default.

Time management by default is neither the most beneficial in terms of work performance nor wellbeing enhancement; thought it has the benefit of keeping things going; their users have little agency to benefit from such a system against organizational time structures and norms that could even have little relation to the real objectives of the organization or the wellbeing of its members. On the other hand, organizations with time management by default have great areas of opportunity to improve performance and lots of low-hanging fruit in terms of wellbeing enhancements. These potential benefits can be realized by transitioning to what we call time management by intention.

Time management by intention is explicitly stated and required, systematic and purposely embedded in the culture of the organization. It comes accompanied by time structures and norms that are not only used to plan and evaluate, but also to receive feedback from the evaluation system and goals setting of the organization and are adaptable and subservient to its core values and the value sought by its objectives in a cycle analog to that of a process of knowledge discovery (Cios et al., 2007b; Fayyad, 1997); which is why such kind of approaches make possible the development of intelligent tools of time management that learn how to best recommend assignations of time use to the members of the organization. Users of time management by intention have high levels of agency to benefit from as, at this point, the only limits are the flow of information across the organization and its processing.

Intelligent time management is the practice defined and proposed by the research team. Intelligent time management is time management by intention aided by a system of data analytics that feeds

AI tools to help members of the organization perform high value actions taking into account all relevant information across the organization and knowledge generated from the processing of this information. Users of intelligent time management may have the highest level of agency to be had in their organization to benefit from it. A comparation with the other two levels of time management can be seen in **Figure 3**. For this kind of powerful tools, wellbeing must be a core criterion.

Time management by default

- Minimal benefits for performance (keeps things going).
- Minimal benefits for wellbeing.
- Low agency to benefit from it.
- Wellbeing factors not considered or not relevant.
- Can go against time structures and norms.

Time management by intention

- Enhaced performance.
- Enhaced wellbeing.
- High agency to benefit from it.
- Considerations of wellbeing optional.
- Comprehesnive process of negotiation of time structures and norms.

Intelligent time management

- Virtuous cycle of enhaced performance and wellbeing.
- Highest agency to be had in the organizaton to benefit from it.
- Consideration of wellbeing required.
- Process of negotiation of time structures and norms aided by AI tools.

Figure 3: *Attributes of different levels of time management systems*

Time management practices are usually taken as a performance enhancing practice. If wellbeing is not a core criterion of an intelligent time management system, the performance criterion will push every member of the organization to their limits in a non-sustainable way, beginning with the people whose individual differences make them benefit the least from time management. The questions become not if people will suffer from burn out, if performance will decrease, if members of the organization will leave, if certain roles will have high turnover… the questions become an issue of when these things will happen. And the only way wellbeing becomes a core criterion of an organizational time management system is if it is first a core value of the organization.

Many organizations only pay lip service to some kind of wellbeing construct for their members and stakeholders. Others relegate wellbeing to the sidelines in activities outside work or consider wellbeing as an individual pursuit that the organization should only encourage mildly. Discounted gym memberships and social integration activities can only go so far if wellbeing is not put front and center in deciding how the core activities of the organizations will be done by their members. Non-integrated approaches to wellbeing and wellness have limited evidence in their favor on wellbeing enhancement, whereas integrated approaches that consider for protecting and promoting worker safety, health and wellbeing, among others factors have growing evidence of their benefits

(Sorensen et al., 2018) and therefore, of their link to performance, particularly if the culture of the organization does not already contains some encouragement to the procurement of wellbeing. On the other hand, effective approaches to wellbeing include injecting it directly into the core activities of the organization through some time management system and this is only possible if wellbeing is also a factor in the time norms and structures of the organization.

At this juncture, we can see the depth of the marriage of time management and wellbeing in organizational settings. There is clear evidence that effective time management enhances performance, advancing more efficiently behaviors that an organization would like to promote because they benefit its goals. Also, there is clear evidence that time management can enhance the wellbeing of the organization's members and that enhanced levels of wellbeing are correlated with further enhancements of performance even without changes in time management, and there is evidence of wellbeing enhancing time management behaviors and results (Aeon & Aguinis, 2017).

Wellbeing, time management and high performance are implicated in a virtuous cycle for which there is clear evidence, even if somewhat muddied by the problems of lack of agency and explicit time norms and structures, as we will see in the next chapter.

A well-designed intelligent time management system with a foundation of a data analytics system, core organizational values aligned with wellbeing, and a consideration of time structures and norms can help members of an organization to perform high value actions that lead to both enhanced performance and wellbeing and the virtuous cycle and high agency that this state implies; as we can see in **Figure 4**.

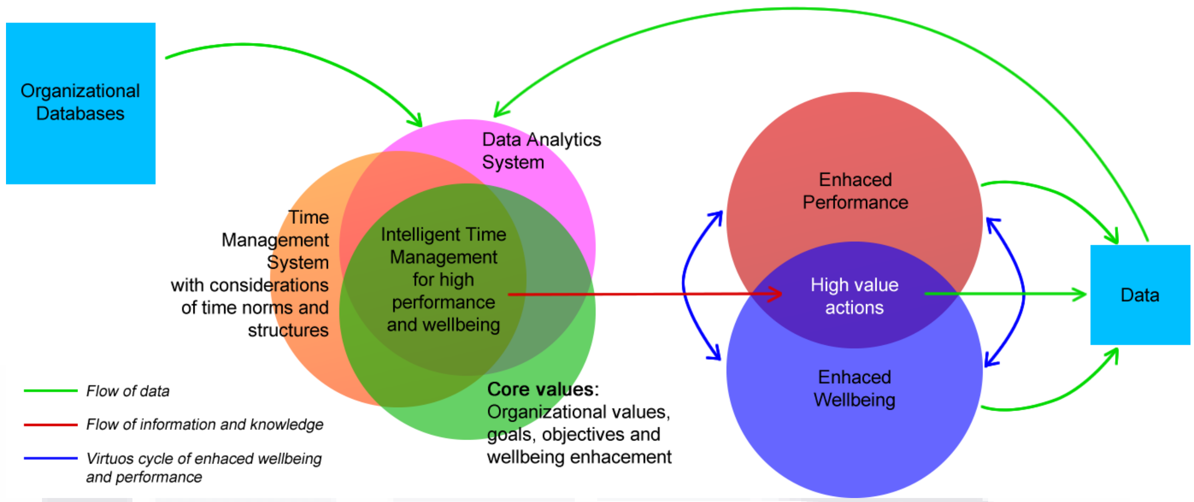


Figure 4: *A diagram for an implementation of an intelligent time management system with wellbeing as core value system; blue arrows represent the virtuous cycle of further enhanced performance leading to further wellbeing by means of high value actions; red arrows represent the flow of information and knowledge from the time management system to the person to facilitate high value actions; green arrows represent flows of data*

This virtuous cycle mentioned above can be realized so long as the complex factors that go into the construction of the organizational culture which will house the time management wellbeing centered system are taken into account and negotiated into a participatory agreement amenable to high levels of agency to take high value action that lead to high levels of performance and wellbeing; which, as we have seen is the most complex task, but can be aided with a data analytics approach and intelligent tools resulting from this approach.

III. Literature review and state of the art

Having presented the concepts and theories on which this dissertation is based, we will present the reviews of the evidence of the linkages between time management, wellbeing, and performance as well as how it fits into the research and experiments done.

III.1. Evidence of time management being linked to better performance

"Time management seems to have more consistent effects on performance defined as behaviors compared to performance defined as results or outcomes" (Aeon & Aguinis, 2017). That is, as with the example of the engineers in the previous section, the implementation of time management was successful in helping some of the engineers to manage their time (by adopting some new behaviors) more efficiently notwithstanding that this resulted in them contravening time norms about staying late at work being implicitly expected of hard-working employees at this organization. Therefore, as Aeon and Aguinis (2017) make it clear, we need the context of organizational time norms and structures to make sense of the results of some of the studies collected in their review, something that is lacking in most of them. Nevertheless, the conclusions and correlations found by the review point in the direction of a positive correlation between time management practices and performance.

The results of this review are in **Table 4**, and we can see that whenever the correlation to performance is sought, we have all kind of results: positive, negative, and absent. But whenever the researchers are looking at the correlation to behaviors, we see positive results.

Table 4: *Review of research on the relationship between time management and performance (Aeon & Aguinis, 2017)*

Year	<u>**Authors**</u> ***Sample*** Research design and measures	Conclusions and **correlation**
1996	<u>**Barling et al.**</u> ***102 salespeople (car sales)*** Self-report questionnaire	Time management alone does not correlate with job performance as measured by objective sales, although time management does interact with achievement striving in predicting sales **(r = 0.32).**
1991	<u>**Britton and Tesser**</u> ***90 undergraduate students***	Time management correlates with academic achievement as measured by GPA **(r = 5 –.10 to .39).**

	Self-report questionnaire (longitudinal)	
2004	**Claessens et al.** ***70 R&D engineers (semiconductor industry)*** Self-report questionnaire	Time management is associated with self-reported job performance **(r = 0.33).**
2014	**Häfner, Oberst, et al.** ***96 undergraduate students*** Experiment (time management intervention)	Time management intervention reduces procrastination **(partial $\boldsymbol{\eta^2 = 0.21}$).**
2010	**Häfner and Stock** *71 employees (trading company)* Experiment (time management training)	Time management training has no impact on performance as assessed by supervisors.
1982	**Hall and Hursch** ***4 participants (university faculty and staff)*** Time management intervention without control group	Time management is associated with an increase in time spent on high-priority tasks and self-rated effectiveness.
2013	**Käser et al.** **196 university students** Experiment	Dedicating uninterrupted time to work on some tasks (i.e. quiet time) leads to lower performance.
2013	**König, Kleinmann, and Höhmann** ***27 managers (financial sector)*** Experiment	Dedicating uninterrupted time to work (i.e. quiet time) leads to higher self-reported job performance **(β = 0.83).**
1994	**Macan**	Time management is not associated with job performance.

	Study 1: 353 employees; study 2: ***341 undergraduate students*** Self-report questionnaire	
1996	<u>**Macan**</u> ***44 employees (social service agency)*** Quasi-experimental field study (time management training)	Time management training does not lead to more time management behaviors and does not increase job performance.
1990	<u>**Macan et al.**</u> ***165 graduate and undergraduate students*** Self-report questionnaire	Time management is associated with higher self-reported performance as measured both by perceptions **(r = 0.32)** and GPA **(r = 0.23).**
2011	<u>**Nonis, Fenner, and Sager**</u> ***201 salespeople (various sectors***) Self-report questionnaire	Time management correlates positively with self-reported job performance **(r = 0.13 to 0.43).**
2005	<u>**Nonis et al.**</u> ***205 MBA students (U.S. and Sri Lanka)*** Self-report questionnaire	Time management is associated with higher = self-reported job performance **(r = 0.06 to 0.26).**
1994	<u>**Orpen**</u> ***52 supervisors (manufacturing sector)*** Experiment (training program)	Time management training increases job performance as assessed by managers' appraisal of participants' activity diaries.
2013	<u>**Rapp et al.**</u> ***212 employees and 41 supervisors (hospitality industry)***	Time management correlates with the influence of helping behavior **(r = 0.16)** on job performance.

	Self-report questionnaire	
1993	<u>**Slaven and Totterdell**</u> ***32 employees (various sectors)*** Time management intervention (no control group)	Time management training is not associated with motivation, commitment, and time spent on high-priority tasks.
1996	<u>**Trueman and Hartley**</u> ***293 university students*** Self-report questionnaire	Time management is associated with overall academic performance **(r = 0.21).**
2003	<u>**Van Eerde**</u> ***37 trainees*** Quasi-experimen	Time management reduces procrastination $\boldsymbol{\eta}^2 = \mathbf{0.10}$
1986	<u>**Woolfolk and Woolfolk**</u> ***81 pre-service teachers (undergraduate seniors)*** Experiment (time management training)	Time management training does not increase performance ratings as assessed by cooperating teachers and supervisors.
2010	<u>**Zampetakis et al**</u>**.** ***186 undergraduate students*** Self-report questionnaire	Time management is positively associated with creativity **(r = 0.48).**

As we can see, in **Table 4**, the results contained in the existing literature linking time management and work performance have a lot of variability. The authors of the revieew argue that this is because time management research should use “time structures and norms as a lens to study time management at the team, organization, and country levels of analysis” (Aeon & Aguinis, 2017) and the studies in **Table 4** did not use this approach. It is relevant to see that the studies that did find positive correlation between time management and performance are mostly those having to do with academic grades and self-reported performance, that is, one in an environment in which forms of evaluation are mostly explicit and another in which the evaluation is subjective. While the second one could be attributed to the subjective perspective, academic grades are objective measures of performance (notwithstanding the differences in how they are calculated in different institutions) that are greatly dependent of certain kind of self-managed behaviors such as keeping a regular schedule of activities such as studying, homework, class attendance, sleep, and so on.

Attention should be called to the study done by Orpen (1994), as Aeon and Aguinis (2017) argue, which is one the few that found a positive correlation between time management training and performance as evaluated by a supervisor. It also was the first experiment to feature a control group to better assess the effects of the treatment. The results of this study found that both the supervisors' assessment of performance and the self-assessment was positively correlated with how much the employees thought that they had benefited from time management training and if the training had met its objectives (Orpen, 1994). Also, we can see that the group who participated in the study "were doing jobs in which the company personnel manager felt that employee performance would be improved by better time management" (Orpen, 1994). In other words, we see that the organizational culture already valued the kind of outcomes expected of time management; and while it is not explicitly mentioned, such culture necessarily includes time structures and norms amenable to such outcomes. With this context in mind, it is less surprising that this is the kind of study that finds positive correlation between time management and performance as evaluated by a third party.

This is not to say that if such context were considered by the other researchers, they would have found positive correlations. What we suggest is that for such studies to find positive correlations more consistently, they would need to be done after the time structures and norms of the groups are amenable to effectively apply time management practices, and that the implementation of these practices must include the consideration and, if necessary, modifications of time structures and norms at the team or organizational level. In other words, inadequately implemented time management systems will more likely not have the expected positive results on performance evaluations by third parties.

Unsurprisingly, Aeon and Aguinis (2017) call for research using the approach of considering time structures and norms to better assess the effects of time management practices on performance.

III.2. Evidence of time management being linked to enhanced wellbeing

In the case of the link between time management and wellbeing, the existing evidence is clearer than with performance. The review's authors suggest that "overall, time management may be useful for well-being enhancement and stress relief" (Aeon & Aguinis, 2017).

The results of Aeon & Aguinis' (2017) review on time management and wellbeing are in **Table 5**.

Table 5: *A selection of research on the relationship between time management and wellbeing (Aeon & Aguinis, 2017)*

Year	Authors *Sample* Research design and measures	Conclusions and **correlation**
1999	**Adams and Jex** **522 working adults/part-time students** Self-report questionnaire	Time management correlates with health **(r = 0.39)** and job satisfaction **(r = 0.27)** indirectly through perceived control of time and a reduction of work–family conflict.
1988	**Bond and Feather** **Undergraduate psychology students (sample 1 5 336; sample 2 5 193; sample 3 5 217)** Self-report questionnaire	Time management is positively associated with better health **(r = 0.27),** a sense of purpose **(r = 0.65),** and optimism **(r = 0.31),** and negatively related to depression **(r = –0.44)**, psychological distress **(r = –0.37),** and anxiety **(r = –0.56).**
2011	**Chang and Nguyen** **111 undergraduate students** Self-report questionnaire	Time management correlates positively with job satisfaction **(r = 0.31)** and psychological well-being **(r = .31).**
2004	**Claessens et al.** **70 R&D engineers** Self-report questionnaire	Time management is associated with job satisfaction **(r = 0.30)** and work strain through perceived control of time **(r = –0.58).**
2010	**Häfner and Stock** **71 employees (trading company)** Experiment (time management training)	Time management training is negatively related to stress and increases perceived control of time.
2014	**Häfner, Stock, et al.** **177 undergraduate students** Experiment (time management intervention)	Time management training reduces perceived stress (**partial** $\boldsymbol{\eta^2 = 0.03}$) and increases perceived control of time (**partial** $\boldsymbol{\eta^2 = 0.03}$).

2015	<u>Häfner et al.</u> **23 undergraduate students** Non-equivalent dependent variable design (time management intervention)	Time management training increases perceived control of time and reduces perceived stress.
1999	<u>Jex and Elacqua</u> **525 full-time employees/ part-time students** Self-report questionnaire	Time management is negatively associated with strain **(r = –0.15 to –0.42).**
2003	<u>Kelly</u> **130 undergraduate students** Self-report questionnaire	Time management is negatively related to worry **(r = –0.21),** although an alternative measure of time management showed no significant correlation **(r = 0.04).**
1992	<u>Lang</u> **96 full-time and part-time employees (taking evening business classes)** Self-report questionnaire	Time management correlates with less anxiety **(r = –0.22)** but not depression and somatic symptoms.
1994	<u>Macan</u> **Study 1: 353 employees (Various organizations); study 2: 341 undergraduate students** Self-report questionnaire	Time management is related to perceived control of time **(r = –0.04 to 0.43),** which in turn relates to increased job satisfaction **(r = 0.29)** and reduced stress **(r = –0.32).**
1996	<u>Macan</u> **44 employees (social service agency)** Quasi-experimental field study (in-house time management training)	Time management training increases perceived control of time and reduces somatic tensions.

1990	**Macan et al.** **165 graduate and undergraduate students** Self-report questionnaire and grade point average	Time management is associated with less role ambiguity **(r = –0.47)** and somatic tension **(r = –0.26),** and with greater job **(r = 0.26)** and life satisfaction **(r = 0.23).**
2000	**Misra and McKean** **249 university students** Self-report questionnaire	Time management correlates with less academic stress **(r = .006 to –0.39).**
2003	**Nonis and Sager** **201 sales representatives** Self-report questionnaire	Time management correlates negatively with **stress (r = –0.19 to –0.32).**
2005	**Nonis, Teng, and Ford** **205 MBA students (U.S. and Sri Lanka)** Self-report questionnaire	Time management correlates positively with job satisfaction **(r = 0.18 to 0.39).**
2005	**Peeters and Rutte** **123 elementary teachers** Self-report questionnaire	Time management correlates negatively with emotional exhaustion in people who have low autonomy and high work demands **(r = –0.17).**
2003	**Van Eerde** **37 trainees** Quasi-experiment	Time management reduces worrying **($\eta^2 = 0.08$)**
2013	**Van Hoye and Lootens** **231 unemployed people** Self-report questionnaire	Time structuring correlates with psychological well-being during unemployment **(r = –0.12 to 0.52).**
1997	**Wanberg, Griffiths, and Gavin** **243 unemployed and employed individuals** Self-report questionnaire (longitudinal)	Time structuring correlates with better mental health among unemployed people **(r = 0.19).**

As we can see in **Table 5**, a big majority of the studies reviewed report positive correlations with positive affects linked to wellbeing, as well as to job and life satisfaction, sense of purpose, optimism and with factors defined as psychological wellbeing and mental health. Also, negative correlations are reported with negative affects linked to lack of wellbeing such as stress, anxiety, depression, strain, somatic symptoms and tensions, and worry. A couple of studies report nonsignificant correlations or no correlations. However, the evidence is so great about the benefits to wellbeing of time management that it has led to this practice being suggested to people with burn out once it has been diagnosed (Peeters & Rutte, 2005).

We should call attention to Peeters and Rutte (2005) experiment with self-reported questionaries from 123 elementary teachers, which found negative correlation between time management and emotional exhaustion in people who have low autonomy and high work demands, both markers of a situation of low agency not conductive to wellbeing as defined in this work. Taking the definition proposed in the previous section of this work as a state of human beings in which their mental and bodily state allows them to act while gaining greater ability to interact with the enviroment we can glimpse that time management may be a useful tool for gaining wellbeing even in circumstances in which agency, and therefore wellbeing, is highly curtailed as the reduced negative affects reported in the study after the introduction of time management techniques indicate a lessening of the process of wellbeing reduction. This is particularly evident in this study because it reports that “high work demands, and low autonomy do result in emotional exhaustion unless one manages one’s time to compensate for this negative effect. The higher the work demands, the stronger the effect of time management” (Peeters & Rutte, 2005), which highlights the already abundant evidence that time management is a viable tool to introduce wellbeing considerations into an organizational setting with many areas of opportunity or so called low-hanging fruit.

III.3. Evidence of wellbeing being linked to better time management

One of the shortcomings of most of the time management research so far, is that causation is not firmly stablished (Claessens et al., 2007), so there is a possibility that some of the evidence presented in **Table 5** can be used to make the case that people with wellbeing do manage and value their time better, or benefit the most from time management training to begin with; and many researchers reviewed admit as much (Aeon & Aguinis, 2017). Given that the definition of wellbeing in this work features prominently a growing ability to interact with the environment, one should consider the possibility that there are cases in which wellbeing comes first, and good time management second, and in fact we argue that both should be sought together for maximum beneficial effect.

Aeon and Aguinis (2017) put forward an example from a study that compared workers paid by the hour an average of 20 cents (USD) by Amazon's online labor system *Mechanical Turk* with participants from a middle-class background, in which evidence was found that the first group valued their time less than the second (Goodman et al., 2013), as it could initially be hypothesized from their willingness to sell it at 20 cents an hour. Also, there is evidence that people who have a sense of their time having a high value, are more likely to engage in good time management practices (DeVoe & Pfeffer, 2007). It could be argued that people with less wellbeing, that is, less ability to interact with their world (and financial scarcity certainly qualifies as less ability to interact with the environment) are less likely to engage in time management because they tend not to value their time as much as people with high levels of such an ability, that is, a state of wellbeing would imply a greater possibility of a person engaging in time management practices and extracting better benefits from it. In other words, high levels of wellbeing correlate with high levels of agency.

This is not to say that time management only offers benefits to those with high levels of wellbeing. We also have evidence that even in cases of diminished ability to interact with the environment, as with cases of low autonomy and high demands works, time management does result in enhanced wellbeing and, in fact, its effects is more pronounced the less autonomy the person has to begin with (Peeters & Rutte, 2005). So, we are in good grounds to act with two twin hypotheses in mind from the evidence presented in the two previous sections:

1) Groups of people with wellbeing are more likely to benefit from and apply time management and benefit from training in it.
2) Groups of people who use time management are more likely to further enhance their wellbeing.

If an organization greatly values structuring its roles and activities to favor its members constantly further their ability to interact with their environment, it should be expected that this organizations not only considers important the professional, intellectual and social growth of its members in order that they should constantly be better in all aspects of their lives including their job activities; not only meeting their expected goals but also finding new ways of accomplishing the organization's goals, and even setting new ones organically. One could reasonable conclude that such an organization would have high levels of wellbeing among its members and that the power of time management systems and tools would be something this organization would find desirable for its members; hence, based on the evidence, they being more likely to apply good practices of time management and the time structures and norms allowing them to extract the benefits of this practices unimpeded, chief among them further enhancement of wellbeing. Therefore, we see that this is where the virtuous cycle may close in a sequence of further wellbeing leading to better use of time, and better use of time leading to further wellbeing, as time-use research already suggest (Gershuny, 2015) as do research in economics and wellbeing (Karacaoglu et al., 2019).

III.4. Evidence of wellbeing being linked to better performance

As for wellbeing resulting in better performance, Warr and Nielsen (2018) review of wellbeing and job performance confirms the expectations of wellbeing being associated with job performance, as after reviewing the literature in this issue they conclude that "the summary of empirical research has revealed a consistently positive but small association between individual workers' wellbeing and performance. Many studies have additionally controlled for potentially distorting demographic and other variables, and a modest cross-sectional association seems to be well established across many work behaviors" (Warr & Nielsen, 2018). The authors also consider the possibility that the found causal relationships might be reciprocal between a variety of constructs of wellbeing and job performance, though they openly admit the causality waters are muddied. However, none of the studies reviewed considered the role agency levels had in the people belonging to the groups studied.

Warr and Nielsen (2018) review collects evidence from a variety of studies done about the relationship between different constructs of wellbeing and job performance. Wellbeing in the evidence presented is mostly concerned with the subjective dimension of wellbeing as measured by cognitive-affective compounds. As we have seen, for the purpose of this work, positive and negative affects are indicators of being in the process of enhancement or decrement of individual wellbeing, so we can use the data provided by the authors of the studies, if not as an equivalent to levels of wellbeing, as a measure of people being in the process of an enhancement or decrement of wellbeing. Also, in some of the studies, wellbeing is defined in a specific context such as:

- *Context free wellbeing: Concerned about life in general.*

    - *Satisfaction with life.*
    - *Global happiness.*
  - *Medium-scope wellbeing:*
    - *Family.*
    - *Health.*
    - *Leisure.*
    - *Other scopes.*
  - *Domain specific wellbeing:*
    - *Job satisfaction.*
    - *Job engagement.* (Warr & Nielsen, 2018)

On the other hand, the studies reviewed also deal with different kinds of definitions of job performance obtained, most usually by third party evaluations and self-description, as objective indicators such as outputs and sales were rarely available. Five kinds of job performance are reviewed:

> ***Task performance,*** *sometimes referred to as 'in-role performance' or 'proficiency', has been of primary interest. Focusing on behaviors which are formally required to meet organizational goals, research has investigated either overall indicators of a person's effectiveness in a job or the summation of separate behavioral assessments.*
>
> ***Specific job behaviors*** *have been examined without setting those in the overall-performance category above. Examples have included work quality, adaptability, innovativeness, proactivity, participation in learning, and technical competence.*
>
> ***Organizational citizenship behavior (OCB),*** *also referred to as 'extra-role'. 'contextual', or 'pro-social' activity, is that which goes beyond formally prescribed job goals ('task performance' above), for example through helping colleagues, guiding new workers, or choosing to take on additional tasks.*
>
> ***Counterproductive work behavior*** *has negative value to an organization, violating accepted conventions, for example through damaging equipment, stealing property, bullying subordinates or other people, avoiding effort, and abusing drugs.*

> ***Workers' absenteeism****, another key negative variable in this area, can create difficulties for team colleagues and for wider organizational success, especially when a person's non-attendance is unexpected or long-lasting.* (Warr & Nielsen, 2018)

We can see in **Table 6**, the results of 64 experiments about the relationship between different constructs of wellbeing and job performance.

Table 6: *Results from research papers into the correlation of different constructs of wellbeing and performance metrics (Aeon & Aguinis, 2017)*

Year	**Authors** Kind of wellbeing researched v.s. kind of job performance	Conclusions and **correlation**
1984	**Petty, McGee, and Cavender** Job Satisfaction[1] v.s. Task performance	Average uncorrected correlations between overall job satisfaction and task performance, **r = 0.23**
1985	**Iaffaldano and Muchinsky** Job Satisfaction[1] v.s. Task performance	Average uncorrected correlations between overall job satisfaction and task performance, **r = 0.25**
2001	**Judge, Thoresen, Bono, and** Job Satisfaction[1] v.s. Task performance	Average uncorrected correlations between overall job satisfaction and task performance, **r = 0.18**
2006	**Harrison, Newman, and Roth** Job Satisfaction[1] v.s. Task performance	Average uncorrected correlations between overall job satisfaction and task performance, **r = 0.18**
2016	**Ayala, Peiró Silla, Tordera, Lorente, and Yeves** Job Satisfaction[1] v.s. Specific job behaviors	Young workers' job satisfaction was correlated **r = 0.41** with their self-reports of innovation at work
1995	**Organ and Ryan** Job Satisfaction[1] v.s. OCB	Identified uncorrected average cross-sectional correlations of **r = 0.23, r = 0.20,** and **r = 0.38** with altruism, generalized contribution and overall citizenship.

2006	**Harrison et al.** Job Satisfaction[1] v.s. OCB	Recorded average correlations of **r = 0.26** (concurrent analyses) and **r= 0.22** (predictive, with satisfaction measured before OCB).
2008	**Edwards et al.** Job Satisfaction[1] v.s. OCB	Found a correlation of **r = 0.15** with supervisor-rated citizenship
2002	**Fisher** Job Satisfaction[1] v.s. OCB	Revealed workers' average job satisfaction to be correlated **r = 0.23** with self-reported citizenship behavior,
2006	**Ilies, Scott, and Judge** Job Satisfaction[1] v.s. OCB	Average satisfaction across three weeks that correlation was **r = 0.52**
2005	**Penney and Spector** Job Satisfaction[1] v.s. Counterproductive work behavior	Job satisfaction has been found to be negatively correlated with counterproductive work behavior **r = -0.33** and **r = -0.25** with self-reports and peer-reports
1997	**Johns** Job Satisfaction[1] v.s. absenteeism	Negative correlations have been found, **r = -0.20** to **r = -0.25**
2006	**Harrison et al.** Job Satisfaction[1] v.s. Absenteeism	Negative correlations have been found**, r = -0.14** with meta-analysis.
2000	**Hardy, Woods, and Wall** Job Satisfaction[1] v.s. Absenteeism	Studied job satisfaction in advance of information about absence frequency; the correlation was found to be **r = -0.23.**
1977	**Mirvis and Lawler** Job Satisfaction[1] v.s. Absenteeism	Focusing only on intrinsic aspects of job satisfaction recorded a correlation of **r = -0.81.**
2008	**Edwards et al.** Job Satisfaction[1] v.s. Task performance	Satisfaction with the work itself and with one's supervisor are significantly associated (around **r = 0.20**) with task performance

1984	**Petty et al.,** Job Satisfaction[1] v.s. Task performance	Satisfaction with the work itself and with one's supervisor are significantly associated (around **r = 0.20**) with task performance
2010	**Rich, Lepine, and Crawford** Job engagement[1] v.s. Task performance	Meta-analysis to report mean correlations of job engagement with task performance of **r = 0.35**
2010	**Rich, Lepine, and Crawford** Job engagement[1] v.s. OCB	Meta-analysis to report mean correlations of job engagement with OCB of **r = 0.35**
2011	**Christian et al.** Job engagement[1] v.s. Task performance	Meta-analysis to report mean correlations of job engagement with task performance of **r = 0.36**
2011	**Christian et al.** Job engagement[1] v.s. OCB	Meta-analysis to report mean correlations of job engagement with OCB of **r = 0.26**
2008	**Halbesleben and Wheeler** Job engagement[1] v.s. Task performance	Found a correlation of **r = 0.32** between job engagement and supervisor-rated task performance
2009	**Xanthopoulou, Bakker, Demerouti, and Schaufeli** Job engagement[1] v.s. Branch income	Among employees in a fast-food company, reported a correlation of **r = 0.34** between day-level job engagement and that day's branch income
2015	**Demerouti, Bakker, and Halbesleben** Job engagement[1] v.s. Task performance	A significant association of task performance with job engagement was found in a broader sample of workers (self-reported)
2015	**Demerouti, Bakker, and Halbesleben** Job engagement[1] v.s. Counterproductive work behavior	A negative weak association of counterproductive work behavior with job engagement was found in a broader sample of workers (self-reported)

2010	**Bakker and Bal** Job engagement[1] v.s. Job performance	Teachers' self-rated performance co-varied with their level of job engagement.
2014	**Gorgievski, Moriano, and Bakker** Job engagement[1] v.s. Specific job behaviors	Among self-employed entrepreneurs observed a correlation of job engagement **r= 0.56** with self-described innovative behavior.
2016	**Djkhuizen, Gorgievski, van Veldhoven, and Schalk** Job engagement[1] v.s. sucess	In a sample of entrepreneurs, job engagement was found to be significantly associated with subjective personal, but not financial, success.
2006	**Taris** Job burnout[2] v.s. Task performance	Reported mean correlations with task performance and task performance of **r = -0.19**
2006	**Taris** Job burnout[2] v.s. OCB	Reported mean correlations with task performance and citizenship behavior of **r = -0.17**
2015	**Demerouti et al.** Job burnout[2] v.s. Task performance	Non-significant association
2010	**Janssen, Lam, &Huang,** Job burnout[2] v.s. Task performance	Significant correlations of **r = -0.13**
2014	**Demerouti, Bakker, & Leiter,** Job burnout[2] v.s. Task performance	Significant correlations of **r = -0.27**
2014	**Mastenbroek et al.,** Job burnout[2] v.s. Task performance	Significant correlations of **r = -0.19** and r = **-0.18** with self-report and ratings by third parties respectivaly
2015	**Petrou, Demerouti, & Schaufeli,**	Significant correlation (across time) **r = -0.28**

	Job burnout[2] v.s. Task performance	
2010	**Janssen et al.** Job burnout[2] v.s. OCB	Obtained **r = -0.19** and **r = -0.23** with two different measures,
2014	**Mastenbroek et al.** Job burnout[2] v.s. OCB	Association was insignificant in the study
2008	**Zelenski,Murphy, and Jenkins** Life satisfaction[3] v.s. Job performance	Found an intercorrelation of **r = 0.25** wit self-rated job performance.
2012	**Erdogan, Bauer, Truxillo, and Mansfield** Life satisfaction[3] v.s. Task performance	Average correlation between life satisfaction and task performance of **r = 0.14**, from unspecified number of publications
2011	**Ford, Cerasoli, Higgins, and Decesare** Life satisfaction[3] v.s. Task performance	Average **r = 0.16** from six studies.
2011	**Ford et al.** Psychological distress[4] v.s. Job performance	Generalized anxiety and depression were on average correlated **r= -0.15** and **r= -0.14** (uncorrected) with measures of job performance.
2000	**Hardy et al.** Psychological distress[4] v.s. Absenteeism	Studied context-free psychological distress in relation to subsequent lost-time absence, finding a correlation **of r =0.25**
2017	**Djkhuizen, Gorgievski, van Veldhoven, and Schalk** Job satisfaction[1], job engagement[1], and life satisfaction[3] v.s. Job-specific factors	Wellbeing of entrepreneurs predicted reports of subjective personal and financial success after an interval of two years; however, the lagged association with Business performance (annual profit, financial turnover, and number of employees) was non-significant.

1986	**<u>Motowidlo,Packard, and Manning</u>** Job related affects[1] v.s. Cognitive/Motivational effectiveness	Feelings which can be viewed along an axis from job-related depression to enthusiasm were significantly correlated (**r = -0.31** and **r = -0.21**) with supervisors' ratings of interpersonal and cognitive/motivational effectiveness, but feelings of anxietywere unrelated to performance in those terms.
1994	**<u>Staw, Sutton, and Pelled</u>** Job related affects[1] v.s. Task performance	Investigated supervisor ratings of task performance in relation to job-related feelings along the depression-enthusiasm axis (which they referred to as 'positive emotion'), finding an overall concurrent correlation of **r = 0.30** and a predictive **r = 0.16** over 18 months.
2001	**<u>Eisenberger, Armeli, Rexwinkel, Lynch, and Rhoades</u>** Job related affects[1] v.s. Task performance	Activated positive affect was found to be correlated **r = 0.16** with supervisor-rated task performance.
1991	**<u>George</u>** Job related affects[1] v.s. OCB	Recorded a correlation of **r = 0.24** with activated positive affect with supervisor-rated altruism
1991	**<u>George</u>** Job related affects[1] v.s. Job performance	Recorded a correlation of **r = 0.10** with activated positive affect with sales
2008	**<u>Zelenski et al.</u>** Job related affects[1] v.s. Job performance	Positive affect was concurrently associated **r = 0.36** with self-rated productivity, but correlations with negative affect were around zero
2002	**<u>Fisher</u>** Job related affects[1] v.s. OCB	Stronger association with positive than with negative affect was also found in analyses of average job feelings across two weeks (**r = 0.48** and **r = -0.01** respectively).
2006	**<u>Ilies et al.</u>** Job related affect[1]s v.s.	Average diary levels of activated positive affect across three weeks were found by Ilies etal. (2006) to be

	OCB	correlated **r = 0.61** with self-reported organizational citizenship behavior.
2009	**Foo, Uy, and Baron** Job related affects[1] v.s. OCB	Examining momentary feelings, self-reported allocation of job effort was found to be correlated on average **r = 0.20** with positive affect but only **r = 0.06** with negative affect.
2009	**Dalal,Lam, Weiss, Welch, and Hulin** Job related affects[1] v.s. OCB, counterproductive behavior	Two diary studies yielded significant concurrent associations between job-relatedaffect and self-reported citizenship and counterproductive behavior, such that positive affect was typicallyrelated to citizenship but not to counterproductive behavior, and negative affect was more often related tocounterproductive activity than to citizenship
2014	**Warr, Bindl, Parker, and Inceoglu** Job related affects[1] v.s. Job performance	Showed that work performance is principally associated with positive job feelings that areactivated,
2014	**Madrid, Patterson, Birdi, Leiva, and Kausel** Job related affects[1] v.s. Job performance	Showed that work performance is principally associated with positive job feelings that areactivated,
1989	**George** Job related affects[1] v.s. Absenteeism	Positive mood at work was found to be correlated -0.28 with subsequent absenteeism, but negative mood was unrelated (**r = -0.03**)
2000	**Hardy et al.** Job related affects[1] v.s. Absenteeism	Feelings of job-related anxiety and depression were correlated **r = 0.30** and **r = 0.25** with subsequent absenteeism
2007	**Tsai, Chen, & Liu** Context free affects[1] v.s. Task performance	Found that the context-free activated positive moodsof insurance sales agents were correlated **r = 0.46** with their supervisors' ratings of task performance and **r = 0.36** with their self-assessments.

2005	**Dalal** Context free affects[1] v.s. OCB	Reviewing organizational citizenship behaviors across 23 previous investigations found average correlations of **r = 0.28** and **r = -0.08** with varied measures of positive affect and negative affect respectively.
2005	**Dalal** Context free affects[1] v.s. OCB	Reviewing counterproductive work behavior, across 23 previous investigations found average correlations of **r = 0.28** and **r= -0.08** with varied measures of positive affect and negative affect, respectively.
2009	**Fritz and Sonnentag** Context free affects[1] v.s. OCB	Asked about proactive work behavior as a function of context-free activated positive and negative affect, finding average correlations of **r = 0.36** and **r = 0.13** respectively.
2017	**Cangiano, Bindl, and Parker** Context free affects[1] v.s. OCB	Significant correlations between proactivity and positive affect (but mixed results for negative affect) were reviewed.
2005	**Amabile, Barsade, Mueller,and Staw** Context free affects[1] v.s. Creativity	Creative contribution at work was modestly but statistically-significantly associated with a day's positive feelings.
2004	**James, Broderson, and Eisenberg** Context free affects[1] v.s. Creativity	Creative contribution at work was modestly but statistically-significantly associated with a day's positive feelings.
[1]Wellbeing in a job context [2]Lack of wellbeing in a job context, characterized as job related emotional exhaustion [3]Context free wellbeing [4]Context free negative wellbeing		

As we can see, every experiment reviewed found a positive correlation between different metrics of job performance and SWB in their positive affects component, and a negative correlation between performance and the negative affects component. So, we see that different components of SWB are relevant in an organizational setting not only because its members, and their families and social groups, will benefit from such an approach, but because the organization itself will benefit from the enhanced wellbeing of its members.

And this virtuous cycle not only gains its momentum from the impulse of healthy organizations to invest in continuous improvement and in expanding their reach and goals, but also from their members' betterment. As new objectives and challenges present, members of an organization need to be better prepared, and to use the experience and knowledge acquired both professionally and academically to meet these challenges. But there is evidence that "people with higher levels of education are likely to exhibit higher expectations with respect to economic and other life circumstances, in which case they will require better circumstances in order to reach a given level of subjective wellbeing compared to someone with less education" (Kristoffersen, 2018) so an organizational approach that takes wellbeing into account will not only result in better performance but will also likely help retain highly prepared members and their contributions as the organization expands its horizons.

III.5. Current AI tools for time management and wellbeing

The use of data analytics in wellbeing applications is currently dominated by prediction tools. These are tools that take datasets from individuals or groups and predict current or future states of components of wellbeing, particularly the affects and the life satisfaction components of SWB, whether in a context free form or related to some specific situation of interest to an organization. There are also applications that predict components of objective wellbeing such as health outcomes or states, mainly used for health risk assessments; and financial wellbeing, mostly used to evaluate credit risks. It should be noticed that while the "risk" language is less frequently used in relationship with wellbeing, the lack of wellbeing may also lead to risks that organization may want to prevent.

Research and measurement of wellbeing is not a new development of the big data era, but prior to the possibility of processing datasets classified as big data research in this topic relied in instruments that explicitly asked for wellbeing data such as questionaries and diaries, which were an effective source of information, but a costly one. Reports of medical data were also used and continue to be important sources of wellbeing information (Bellet & Frijters, 2019). Currently, most wellbeing research and applications rely on data produced by users of social media because this source of data is particularly easy and cheap to retrieve for researchers and developers, but other sources of data are becoming popular, such as data from fitness trackers, sensors and from surveillance technology software (Axtell et al., 2019). Data from sensors and fitness trackers are sometimes used to complement analysis of SWB, but are also used to predict objective health states. Some tools require the generation of specific datasets from the population to analyze and others require access to existing organizational databases.

Many applications are already using a comprehensive approach to wellbeing incorporating both objective and subjective components of wellbeing into a single construct or objectives list, but with the exception of some tools and models for public policy making (Karacaoglu et al., 2019), none of these approaches have used a wellbeing construct directly and explicitly defined by the ability of people to interact with their environment, as the definition presented in this work does.

On the other hand, applications of time management with data analytics are far less common and limited to shared database for calendars in office and project management software without much in terms of predictive tools beyond suggestion to autocomplete fields in elements in a calendar, appointment reminder, or task list.

III.5.1. Subjective wellbeing applications

Most current applications of affect prediction or sentiment analysis keep using data from social media because compared to other kinds of data, the extraction of the affects component of subjective wellbeing information from the social media data is easier, or even trivial, given that most of the posts from users of social media are about sharing emotional states with friends and family, sometimes not only with words but also with graphic depictions of emotions, such as *emoticons*. Some of the best results about affect prediction are obtained from such data (Bellet & Frijters, 2019).

On the other hand, life satisfaction regressions have long been bedeviled by low predictability from socioeconomic, time use, and demographic factors, and even to reach levels of r^2 as high as 0.15, it is necessary to add subjective indicators of physical and mental health, with the later explaining more variation in life satisfaction than physical health (Clark et al., 2008)

As we can see in **Table 7**, even using big data sources such as Facebook or X (formerly known as Twitter) the predictability of life satisfaction is poor, except when using word searches from Google Search and averages, after a process of binning, of sizes of personal network. This kind of data is known as digital footprints or digital exhaust. Digital footprints can be used to predict not only life satisfaction but also other "inner lives" measures such as interpersonal trust and reciprocity, known drivers of subjective wellbeing (Bellet & Frijters, 2019).

Table 7: *Review of R^2 coefficients across main studies that have tried to predict survey responses to life satisfaction or happiness questions from alternative Big Data sources; the information collected is extracted from digital footprints left by individuals when they go online or engage with social media networks (Bellet & Frijters, 2019)*

Authors (year)	SWB measure	Big data measure	Big data source	Unit of analysis	R^2
		Individual level predictions			
Collins et al. (2015)	Life satisfaction	Status updates	Facebook	Facebook users	0.02
Kosinski et al. (2013)	Life satisfaction	Type of Facebook pages liked	Facebook	Facebook users	0.028
Liu et al. (2015)	Life satisfaction	Status updates (positive emotions)	Facebook	Facebook users	0.003
Liu et al. (2015)	Life satisfaction	Status updates (negative emotions)	Facebook	Facebook users	0.026
Schwartz et al. (2016)	Life satisfaction	Status updates (topics, lexica)	Facebook	Facebook users	0.09
		Aggregate level predictions			
Algan et al. (2016)	Life satisfaction	Word searches	Google	Trends US weekly time series	0.760

Algan et al. (2016)	Happiness	Word searches	Google	Trends US weekly time series	0.328
Collins et al. (2015)	Life satisfaction	Average size of personal network	Facebook	LS bins	0.715
Collins et al. (2015)	Life satisfaction	Average number of status updates	Facebook	LS bins	0.096
Collins et al. (2015)	Life satisfaction	Average number of photo tags	Facebook	LS bins	0.348
Hills et al. (2017)	Life satisfaction	Words	Google Books	Yearly panel of 5 countries	0.25
Schwartz et al. (2013)	Life satisfaction	Topics and lexica from tweets	X (Formerly known as Twitter)	US counties	0.094

Digital footprint data can be complemented with data from sensors that tell us how someone writes, walks, looks, smells, touches or sounds (Bellet & Frijters, 2019). An all-encompassing approach in which everyone in a group is equipped with sensors to see how they react to situations and to one another has been successfully advanced by the MIT Media Lab in an experiment with young children in the spectrum of autism, with diagnosis of Attention deficit hyperactivity disorder (ADHD) or other sensory challenges, but the researchers expect their findings and results to translate to other areas and industries in order to answer questions such as "What parts of flying an airline create anxiety, which designers may not even be aware of? How can buying a new cellphone be more relaxing? Any service where regulating a person's arousal has importance can benefit from observing individual's [electrodermal activity or] EDA in a natural setting" (Hedman et al., 2012).

One of the most recent advances in this technology is facial recognition, that can not only identify people but recognize their emotional state. Software based in artificial neural networks such as FaceReader can predict emotions such as anger or happiness with levels of accuracy above 91% in both individuals and groups (Bijlstra & Dotsch, 2012). It should be noted that while some of these tools can use measurements of objective wellbeing to predict subjective wellbeing components, they remain in the realm of SWB tools.

Data from digital footprints can be complemented with psychometric data to obtain higher correlations between different kinds of personality and life satisfaction, however this kind of data must be obtained from properly designed instruments and linked to the digital footprints in order to be more valuable (Collins et al., 2015), as it is shown it **Table 8**, where this data show strong correlations with life satisfaction as opposed to affects obtained from linguistic features from posts on social media.

Table 8*: Pearson correlation between personality type data from the mypersonality.org project and affects as expressed in digital footprints from Facebook thought linguistic features analysis.* (Collins et al., 2015)

Personality feature	R (Pearson)	Sample size
Agreeableness	0.988	86073
Conscientiousness	0.986	86073
Extraversion	0.997	86073
Neuroticism	-0.998	86073
Openness	0.901	86073
Anger	-0.105	3505
Negative emotion	-0.106	3505
Swear	-0.148	3505

As we have mentioned before, these technological tools are very powerful and require careful design and use. While they hold great promise for the development of organizational wellbeing enhancing systems that can include time management, and therefore, influence the core activities of the organization; they can also be used to violate privacy and rights of individuals and generate distrust and guardedness across an organization, which would harm wellbeing levels.

Current commercial applications of subjective wellbeing prediction are divided into two target groups: customers and employees, with both including potential ones.

1. **Prediction of consumer wellbeing:** Platforms that use the data generated by the consumer engagement with the company and sometimes social media data available about the consumers, to predict their level of consumption related wellbeing, or consumer satisfaction and enable the relevant departments to prevent negative experiences for the client, identify products or services that generate high levels of satisfaction, identify new markets opportunities, design marketing campaigns, and in general enhance consumer experience and identify further opportunities to extract value from the relationship between consumer and organization. Examples include platforms such as piHAPPINESS, Zonka Feedback, PxidaCX, amoCRM, FreshDesk, Pipedrive and many other customer satisfaction management and customer relationship management platforms that have incorporated a data analytics approach (Capterra, 2019a). There are also consumer applications sold for personal use, such as software that can predict mood through the data of fitness trackers and sensors.

2. **Prediction of employee wellbeing:** Platforms that use the data generated by employees during their workday to either predict emotional states or general satisfaction, or to detect what situations cause negative subjective experiences that are linked to poor job performance, as seen in the data in **Table 6**. Most of the tools in this category implemented existing human resources management paradigms and added to them the wellbeing data into a greater set of metrics and functions. Some most modern applications use data generated from sensors and fitness trackers, and some enterprises have deployed full-blown surveillance software that generates data about every moment using company computers and systems. Examples include many employee engagement and employee wellbeing platforms such as Burner, GoVida, LifeWorks, Jalapeno, Engagedly, Bizneo HR and many other platforms that have incorporate a data analytics approach (Capterra, 2020). Examples of surveillance software includes Isaak, TeraMind, TimeDoctor, and VeriClock (Capterra, 2019b).

Both approaches, customer, and employee wellbeing, can use public data from social media, particularly those directed at costumers whereas those directed at employees mostly depend on data generated by the existing tools with which consumers and employees interact during the sale or provision of a product or service. Both consumer and employee approaches are sometimes supplemented by instruments specifically designed to harvest personality data, such as psychometric tests for employees or surveys for clients. More comprehensive approaches incorporating data from sensors or fitness trackers are newer and less frequently deployed, and surveillance software is mostly used with employees working remotely for companies, but such approaches are already becoming more commonly used to monitor employees in situ; in both cases the surveillance itself might be an obstacle to employee wellbeing.

Employee wellbeing and engagement platforms are more focused in the core activities of the organization, but ultimately just provide information that is not immediately actionable but can nonetheless be considered for beneficial changes in these activities, and to the design or redesign of new instances of time norms and structures, as well as creating a focus on human wellbeing.

III.5.2. Objective wellbeing applications

While most of the data analytics research and applications into wellbeing are done around SWB, applications about objective wellbeing, particularly in its health and financial components have begun to take advantage of a data analytics or AI approach. Financial institutions have long been interested in tools to calculate or predict credit worthiness of clients, whereas insurance companies, health care providers and employers have used health risk assessments for decades. Big data and data analytics approaches have made it possible to simplify the process of assessment of health and financial risks by predicting these components of objective wellbeing with greater power than with more traditional methods such as questionaries.

For the health component of objective wellbeing, there are three main sources of data: fitness trackers, sensors, and medical reports; with the last one being the easier to interpret and extract information from (thought this process never rises to the level of trivial as it does with SWB and some data from social media), but it is usually a highly restricted source of data given legal protections and privacy concerns. Applications that use medical reports are mostly restricted to use in a medical setting by health care providers mostly because the procedures to protect this information can be onerous to many organizations, and most people are unwilling to hand it over to entities unrelated to health care. Some examples include the analysis of images taken by X-rays, ultrasound, or magnetic resonance imaging to help in the diagnosis of some conditions. Also, analysis of patients' medical records is being done to help medical doctors to ascertain prognosis and diagnosis (Bainbridge, 2019).

The other two sources of data, fitness trackers and sensors, are being used to develop applications to predict objective wellbeing such as cardiac health or risk for type 2 diabetes (Li et al., 2017) and even detection of illnesses such as influenza, COVID-19 and other fast spreading viral infections even before noticeable symptoms appear; examples of this are fitness trackers such as FitBit and Apple Watch (Goode, 2020). Also, the health insurance industry has begun to use this data to ascertain health risks and offer people with healthy habits less costly plans. An example of this is Health IQ (Health IQ, 2020).

As for the financial component of objective wellbeing, banks, credit rating agencies and other financial institution have long used financial data and buying habits to predict future financial states and credit worthiness of people whose information they have been granted access to. Data from use of debit and credit cards have long been a big data source for this kind of institution, but more and more they are using other kinds of big data such as posts from social media which can include favorite products, stores and brands, lifestyle data, travel preferences and even indicate creeping financial risks (Skyrius et al., 2018).

III.5.3. Comprehensive wellbeing applications

Some applications using the data analytics approach are already using a comprehensive approach incorporating both subjective and objective measures of wellbeing and using highly developed wellbeing constructs. Two such applications for predictive wellbeing in organizational settings are the Gallup-Healthways Well-Being 5 and Intel Covalence.

Given the ample spectrum of wellbeing components that these kinds of tools analyze, they are of greater relevance to an action-centered approach that considers wellbeing as enhancing the ability of people to interact with their world. However, neither of the platforms mentioned uses this kind of wellbeing construct explicitly though we can see how it is a very compatible approach.

The Gallup-Healthways Well-Being 5 platform uses a wellbeing construct divided in 5 elements of wellbeing based on research by both Gallup and Healthways. The elements of the construct of wellbeing for this platform are:

***Purpose:** Liking what you do each day and being motivated to achieve your goals.*

Social: *Having supportive relationships and love in your life.*

Financial: *Managing your economic life to reduce stress and increase security.*

Community: *Liking where you live, feeling safe and having pride in your community.*

Physical: *Having good health and enough energy to get things done daily.* (Healthways, 2015)

As we can see, we can see some similarities with the more expansive OECD Better Life Index, but also each of these components can be summed up as: in this dimension of life, having what is needed to better interact with the enviroment. Given that this platform's main appeal is lowering organizational healthcare costs, it is interesting to consider that their research has led them to implicitly take an approach that is compatible with the definition of agency-based wellbeing posited in this work, even as the main benefit sold to the organizations is to lower costs, an important goal that is not in opposition to enhancing wellbeing for employes, or at least reducing psychosocial risks, both part of ESG corporate evaluations that most world class companies listed on stock exchanges must undergo (Becchetti et al., 2022; Bradley, 2021), apart from national regulations in particulars countries such as the NOM-035 in Mexico (Norma Oficial Mexicana NOM-035-STPS-2018, 2018).

The Gallup-Healthways Well-Being 5 platform harvests the data from specially designed surveys and there lies its disadvantage as it is a free-standing instrument that must be incorporated into the organization and doesn't take much advantage of data the organization might already have. The approach from the Well-Being 5 system can be compared to a wellbeing centered business intelligence system, as it results are presented in a dashboard setting to provide information that can enable organizations to "improve individual and organizational performance and lower healthcare costs [and] lower the likelihood of hospital admissions, readmissions, and members' healthcare costs" (Healthways, 2015).

As we can see, this system intends to replace standard health risk assessments not only by performing the same functions more efficiently and lest costly, but by incorporating data that is usually not used by these other assessments because it is not medical information properly, such as social, purpose and financial data, but that has been shown to have a direct correlation with health status and outcomes, and also provide wellbeing information that can lead to better job performance, as we have seen there is evidence for this correlation in the data in **Table 4**.

On the other hand, Intel Covalence is a predictive analytics platform whose purpose is "to improve employee wellness by taking advantage of both external and internal preexisting data in organizations and the data produced by Internet of Things (IoT) devices" (Intel Corporation, 2016), as we can see in **Figure 5**.

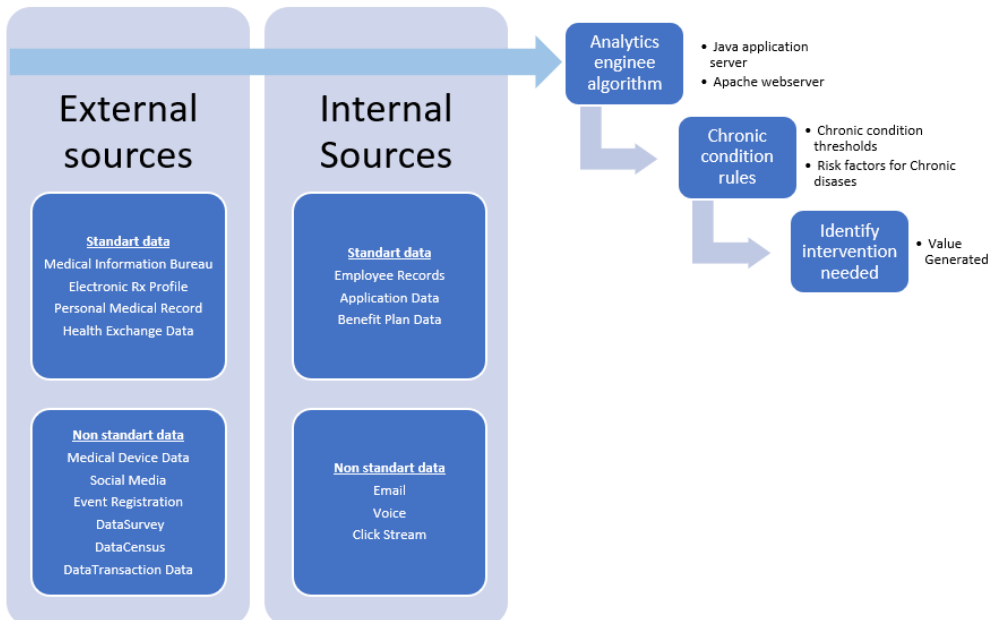


***Figure 5**: Diagram of the Intel Covalence system which is described by the Intel Corporation (2016) as advanced analytics that supports the detection of chronic conditions; automated detection of the onset of chronic conditions driven by data and augmented by clinical counseling can lead to population members better managing their health*

The platform uses various data streams: activity, heart rate, sleep, temperature, demographics, perspiration, blood pressure, blood glucose, body composition, weight, education, perception, dental health, weather, sentiment, claims data, and social media. It relies on different components such as wearable devices, smartphones, IoT devices and gateways, a secure cloud or on premises data lake and an analytics engine. It obtains its data from both existing sources such as policy admins and claims and customer relationship management systems (CRM); and potentially new ones such wearables, sensors, smartphones, social media, scales, blood pressure cuffs, glucometers, and surveys (Intel Corporation, 2016). Its data collection and processing and analytics system incorporates:

- **Data collection and processing**
  - Predictive analytics models
- **Analytics system**
  - Health risk analytics

- Machine learning
- Reporting and visualization
- Data integration
- Data quality
- Data security
- Cohort analytics
- Claim analytics
- Engagement management
- Data visualization and exploration
- Plan churn analytics

The main reported benefits from the platform are for both the employees that benefit from longer, healthier lives and reduced insurance premiums; and for employers that get more productive employees and lower insurance premiums. Other benefits include increased morale and productivity, early interventions, reduced insurance plan churn and new health care insights.

As we can see, both of this highly comprehensive platforms, Intel's and Gallup-Healthways's, have as a main point of selling the health component of objective wellbeing, and use the evidence-based link between physical health and subjective wellbeing to enhance this last one. Both platforms are standalone platforms that can only read existing databases, for example from customer relationship management systems, to get data and do not embed into the core functions of the organization, much less time management practices. This kind of approach is comprehensive, but ultimately is one of wellbeing management from the sidelines. The research teams' review found no publicly available wellbeing platform from a data analytics or AI perspective that targeted students specifically, but many of the tools reviewed can be adapted to students needs, particularly if the intention is to teach them to use them or letting them know these tools before entering the workforce.

III.5.4. Time management applications

Time management can be seen as an algorithm for the optimization of a function constrained by a set of parameters. Under this approach, we must answer questions such as:

- What is the value, or values, to be optimized?
- What are the constraints and how do we choose their parameters?

And so, every application of time management can be modeled in this way. For example, a simple to-do list for the week is a function that is optimized by having all the activities on the list marked as done, and it is constrained by the time available during the week. A more sophisticated to-do list would ask for further data about the tasks such as the time required, delivery date and the importance of each task. In this way a simple weighted ordering algorithm could be used to suggest first the tasks accomplished in the least time or the most important ones. In these instances, there is evidence that both approaches have certain advantages and detriments in terms of behavioral results. For example, accomplishing a multitude of easy tasks in a short period of time can be motivating at first, on the other hand, accomplishing the most important tasks first could be a more efficient use of time long term (Malkoc & Tonietto, 2019), so the approach must be chosen with the individual differences of the user in mind.

Suppose that an intelligent to-do list app asks if some tasks require a simultaneous team effort with a set of people. Such an application could verify the free time slots of every person required in a calendar database and use the parameters of each person to calculate an optimal time for the task, constrained by parameters taken from time structures and norms of the organization. Given enough data generated by these kinds of processes, eventually the application could incorporate a predictive system in which given a certain set of tasks, the occurrence of a task related to these could be predicted in advance and some periods set apart in the calendars of relevant people for it even before the actual need arises. In time, and given some feedback from its users, the application could take a more prescriptive role and suggest or even pre-program sets of tasks in the order and time allocation which is predicted to have an optimal result. At this point, our to-do list would be a very powerful tool of time management interacting with other tools such as calendars and project management apps. And this is just an example of one of the simplest time management tools aided by AI.

Given big data availability, an intelligent time management application could produce personal time profiles that take into account time preferences and the process of time decision making of each individual in order to preserve wellbeing or, if possible, enhance it. This kind of system would prescribe time allocations for the day that consider factors as the most productive times of each member of the organization, team tasks that can be done asynchronously, and psychological and behavioral factors in order to allow each task to be done at the best time possible and minimize behaviors such as delay and procrastination. If an action-based definition of wellbeing is part of this tool, a participatory process of objectives and goals setting would be in place in order to maximize the effectivity of the harmonization process of time allocations of all members of the organization in a way that wellbeing is enhanced, and a maximum amount of agency to manage their time is had by everyone participating on this process in the organization. However, an application of this kind has yet to be developed, and a first step in this direction is one of the main goals of this work: to ascertain the accuracy of these kind of tools in terms of SWB variables linked to time use.

Currently the market for time management applications is just as scattered as the study field is. There are applications for every need imaginable, from classics such as task lists, to-do lists, and calendars, to mind mapping applications and those that automatically recover time-use data and apply an analytics approach to inform users how their time is spent. Most of these applications can be used on a smartphone or as a webapp. There are more comprehensive approaches such as applications for project management and software suites for offices. It is inside this kind of application that intelligent tools are beginning to appear in the form of suggestions of times, places, and sets of people that have been included several times in certain kinds of tasks or appointments. While shared calendars, task lists, and appointments with degrees of importance are already included in these kinds of tools, even comprehensive suites of applications such as Microsoft 365, Google Apps, or Monday.com still lack predictive tools for activities and objectives instead of just to prefill fields for them and they also lack prescriptive tools in general. No application of time management takes wellbeing into account except for those incorporated in the applications mentioned in the last section that are explicitly about wellbeing or wellness. Nonetheless, these applications can hardly fulfill the need for a comprehensive time management tool in an organizational setting; and when they incorporate AI or a data analytics approach, the main benefits or the insights obtained are for the developers of the application and only indirectly for the user if the developers engage in a data-driven redesign of the application.

However, even as the market of time management applications is very rich, most applications are not intelligent and have no use for a data analytics approach to benefit the user but are rather ways to harvest data to better sell the users other products and services, or to sell the data itself to other entities. Most applications that have artificial intelligence incorporated that are of use to the actual users are embedded in suites of organizational management tools and project management. There is no existing time management application that uses tools of artificial intelligence and a criterion of wellbeing to manage core activities in an organizational setting.

III.6. Opportunities and risks of wellbeing data analytics

It can be said that "there is urgent need to examine the impact of wellbeing analytics on organizational behavior and social systems, as well as the legal perspectives, as otherwise there is the danger that such systems could be applied in a way that does not adequately respect the employees' best interests" (Axtell et al., 2019), and Bellet and Frijters (2019) point out that the legal ramifications of such systems can even appear in the form of legal liabilities for organizations that, thanks to these systems, can know in advance about dangers to the physical and mental wellbeing in the workplace that they would otherwise not know about.

Axtell (2019) presents two scenarios, an utopia in which "there is 'co-ownership' between employee and employer relating to the collection, interpretation and use of data, existing within a supportive, caring and trusting culture, where there is a genuine concern for employee wellbeing" and a dystopia in which "there is a lack of trust and openness, and where the initiative is driven purely by organizational needs and not the best interests of employees" and posits that the organizational practice may be "somewhere between these two" (Axtell et al., 2019). This is where a proper definition of wellbeing can be both useful and illuminating. It is not only odd but paradoxical that a system designed or used to measure and predict wellbeing could be used to harm the interest of the very people whose wellbeing is of interest to the organization. But this is only so if wellbeing is merely meant to include its subjective and objective components without a proper orientation to human nature. Let us restate the definition of wellbeing posited in this work:

Wellbeing is the state of human beings in which their mental and bodily state allows them to act while gaining greater ability to interact with their environment.

When a wellbeing centered application, whether it uses data analytics or not, applies a definition of wellbeing set on enhancing the ability of members of an organization to interact with their environment, and therefore to have greater agency to engage in their activities inside and outside the organization, this application necessarily needs to be designed with the active participation and deliberation of all the stakeholders, where concerns about privacy and legitimate uses of the information are not afterthoughts but constituent parts of the foundations of the application, its data governance, architecture, and security. So do the processes of collecting and generating data, and the questions of which actions are worth keeping a tab on, and which would be too intrusive to monitor and would thus generate undesirable effects if measured.

In other words, having a greater ability to interact with the environment can begin by having a greater ability to engage with the work environment (or the academic environment), aided by the intelligent tools with access to broad datasets from the daily activities and patterns or the members of the organization, but not with a primary intention of surveillance. Rather, the primary intention of such systems should be to help bend situation and circumstance to help members of an organization realize their objectives and goals and not bend members of the organization to situation and circumstance. This procedure, given enough data and adequate processing can be done in harmony with the organization’s values and objectives as a whole while furthering its members’ own agency and therefore their own wellbeing capability, which as we have seen in the data in **Table 6** has a correlation with job performance, and other studies have found for students performance (Bücker et al., 2018), even as already shown, most studies and applications are done for and from an employee perspective.

These same opportunities and dangers apply also to objective wellbeing applications, and therefore, to comprehensive wellbeing applications. However, legal restrictions on the use of medical information and health assessments in most countries already protect people from having this information made public or used to discriminate against them. However, the dangers of misuse of this kind of information in an organization remain, and this makes necessary policies about the collection, use and management of information regarding both objective and subjective wellbeing of members.

III.7. Regulatory and corporate criteria national and international implications

Wellbeing applications in organizational settings have another role apart from safeguarding or enhancing the wellbeing levels of the organization’s members: risk prevention. Risk prevention has personal implications for the organization’s members and legal implications for the organization itself and, therefore, possible financial impact if these risks materialize.

Also, national regulations must be complied with. For example, a national regulation that seeks to identify, analyze, and prevent risk factors, particularly psychosocial risk factors, such as the Mexican NOM-035. This regulation not only focuses on these risk factors but also “promotes a favorable organizational environment in the workplace” (Norma Oficial Mexicana NOM-035-STPS-2018, 2018) and has an analogous regulation in many countries, including the United States and those belonging to the European Union.

Organizations that are national in scope must obey the regulations of their respective countries. In contrast, those that operate in many countries usually adhere to the most stringent one in an effort for uniformity unless this regulation is particularly onerous. However, given the state of the research and application in the field of wellbeing and time management, implementing intelligent tools with a wellbeing criterion would serve both national and international organizations very well in their efforts to comply with regulations that impact wellbeing and its risk factors.

Also, as already mentioned, many enterprises listed in stock exchanges must undergo ESG corporate evaluations, which include a social component that incorporates the prevention of risks that impact the wellbeing of members of the organization and other stakeholders (Becchetti et al., 2022; Bradley, 2021). While these risks can have a legal and financial impact on their own, a lack of risk management itself can have an impact on the ESG evaluations, which can result in an impact on the stock prices of the enterprises and even their delisting in certain indexes, both situations that can have also legal implications.

Given these risks, the use of intelligent time management tools in an organizational setting, whether a business, governmental, non-profit, or academic one, is dependent on these tools accuracy and capacity to offer recommendations that are correct for their objectives, and this where the question of whether these tools can have an appropriate level of accuracy in predicting wellbeing levels arises. Only if they have meaningful levels of accuracy can these tools offer appropriate recommendations and courses of action, hence the relevance of the research questions in this dissertation for organizations in a business or academic context.

IV. Intelligent algorithms used

Artificial intelligence (AI) is not exhausted by machine learning (ML) methods and applications; however, most intelligent algorithms used in our research belong to the ML subset of AI. Nonetheless, we present both ML and non-ML approaches and how their use is justified according to the different characteristics that make them useful for the different applications that will be presented in subsequent chapters. In the ML area, we introduce mainly classification algorithms, used later to classify users and respondents into different levels of SWB, and as part of ML-based automatic feature selection strategies for AI-aided instrument measurement design.

We also use non-ML approaches such as optimization and genetic algorithms to find better allocations of time use. This chapter also includes description of the cross-validation regimes and performance metrics used as well as some basic information about AI and ML.

IV.1. Artificial Intelligence and Machine learning

While AI has no generally agreed definition, we can begin to characterize it as a field of computer science, particularly "the field in computer science that is concerned with the automation of intelligence and the enablement of machines to achieve complex tasks in complex environments" (Batarseh, 2018). In the context of its applications, it can be considered "the effort to automate the intellectual tasks that humans normally perform" (Chollet, 2018). AI is an enormous field, and it is home to one of the most highly valued disciplines of the first twenty years of the XXI century: ML.

As we can see, ML is the automation or emulation of intelligence's learning ability. In other words, the discipline of ML is concerned with creating effective and efficient machines that can learn and have many purposes. A machine in this context can be a logical and mathematical construct implemented in a computer, even as the word *machine* prompts imagery of robots. From its focus on learning, ML can also be defined in many ways.

IV.1.1. Definition of machine learning

As a discipline, ML is the "field of computer science that studies algorithms and techniques for automating solutions to complex problems that are hard to program using conventional programming methods" (Rebala et al., 2019). In the context of the field of AI, "[ML is a] term used in the artificial intelligence community to indicate automated improvement based on experience or empirical data in accomplishing a given task such as optimizing an objective function" (Gass & Fu, 2013). Notice the emphasis on improvement based on experience, as it is central to implementing ML algorithms, hence the importance of clarifying what is it that is going to get improved, and how it will be measured, before implementing an ML solution, as has been argued in previous chapters.

But where is the learning part? Gass and Fu (2013) tell us that the machine encounters empirical data and gains experience. However, it is essential to add that the learning is kept in the machine, which is modified by the learning process. We can say that "a machine learns whenever it changes its structure, program, or data (based on its inputs or in response to external information) in such a manner that its expected future performance improves" (Nilsson, 1997b). In other words, a machine that has learned something is already a different machine than it was before undergoing the learning process. Moreover, it is uncontroversial to say that intelligent beings, including animals and humans, are not the same after they have learned something, as they have assimilated important information about their environment and keep this information in complex structures and relationships between cells and other components of their nervous systems.

Another viewpoint is not to see the ML models as changing as they learn, as their representation of knowledge improves, but to see ML as a search for the most useful representation of a piece of information of relevance to solving some problem or performing a task. Chollet gives us a more succinct definition of ML as a search mechanism, a search "for useful representations of some input data, within a predefined space of possibilities, using guidance from a feedback signal" (Chollet, 2018). In this perspective, we see the intimate interdependency of ML with other non-ML parts of AI or of Operations Research (OR), as there is a search algorithm at the heart of practically all ML methods.

Given this information, it is not surprising that the discipline of ML is focused on two interrelated questions: "How can one construct computer systems that automatically improve through experience? and what are the fundamental statistical computational-information-theoretic laws that govern all learning systems, including computers, humans, and organizations?" (Jordan & Mitchell, 2015). Both questions are highly relevant to many fields of science and engineering that, given the explosion of data available to them, need an ML approach to find the patterns in the data that can prompt further inquiry and scientific discovery or a successful application such as voice recognition, computer vision, or large language models (LLM). Also, ML research and applications have high-stakes roles in studying human intelligence as well as organizations as entities that generate data and transform it into knowledge, learn it, and use it to accomplish their objectives better, which is analogic to the definition of human wellbeing seen in the previous chapter.

As mentioned, an ML algorithm can be viewed as “searching through a large space of candidate programs, guided by training experience, to find a program that optimizes the performance metric” (Jordan & Mitchell, 2015). Therefore, we can see this as opposed to the non-ML AI approaches in which a program has to be expressly developed to solve a problem optimally or as near to it as feasible or practical. Sometimes the task of the researchers or engineers is to reformulate the problem at hand as the problem that the non-ML program they want to use is designed to solve.

ML models are made of various parts. They have a representation of the candidate programs to solve the problem at hand, a search or optimization algorithm to search through those candidate programs, and a way to evaluate how well a program performs. Sometimes the ML paradigm or learning regime is an integral part of an ML model.

ML programs are different from one another in the way in which they represent the candidate programs: decision trees, well defined mathematical functions, or general programming languages constructs; they also differ by how they comb through this search space of possible programs. ML programs can use for this search optimization algorithms that guarantee convergence or are even exhaustive, or heuristic and metaheuristic search methods. ML programs also differ by the methods of learning or ML paradigms of which there are supervised, non-supervised and reinforcement methods. A diagram of the variety of ML programs' features can be seen in **Figure 6**. Other ML approaches emerge from combinations across the three main ML paradigms, such as semi-supervised learning, discriminative training, active learning, and causal modeling (Jordan & Mitchell, 2015). We can see that, while optimization methods and heuristics such as evolutionary search can be considered non-ML AI methods, they are at the heart of ML ones; and that the kind of data available can define the kind of ML approach to take.

Representation of the program	Search algorithm	Method of learning
• Decision trees • Functions • General programming lenguages	• Optimization algorithms • Evolutionary algorithms	• Supervised • Non-supervised • Reinforcement • Combined

Figure 6: *Features of ML programs*

ML programs differentiate themselves from all others in the AI field because of their capacity to learn and change even as they use these very non-ML approaches for their own internal affairs. To gain more clarity about the ML field, we will review what is just on, or beyond, the frontier of ML and what tools and concepts have applications in both the ML world and outside it.

The main methods of learning or learning paradigms can be explained as follows with an example of classification:

Supervised learning: The ML model is fed labeled data. That is, a pair of data points (X, y) in which X can be a vector with any number of usually numeric features, and y is the category or label. For example, in equation (1) we have a dataset in which X is a two-feature vector that contains the weight in kilograms and the height in centimeters of a person, and y is a binary variable if the person has been diagnosed with a medical condition.

$$\left[\left((88,157),1\right),\left((112,189),0\right),\left((95,170),1\right),\ldots\right] \qquad (1)$$

Non-supervised: The ML model is fed uncategorized data. That is, only instances of vectors X with any number of usually numeric features. For example, just the X vectors of the previous example, as shown in equation (2).

$$[(88,157),(112,189),(95,170),\ldots] \qquad (2)$$

Reinforcement: In this learning paradigm, interaction happens not with data but with environments in which many parameters come into play to learn and act accordingly to meet a goal. In most reinforcement learning, the concepts o of positive feedback (or rewards) and negative ones come into play (Nandy & Biswas, 2018). For example, if we wanted a robot to exit a labyrinth, we could classify moving increasing distances without hitting a wall as positive feedback and hitting walls as negative.

IV.1.2. Non-machine learning AI

As we have implied in the previous section, not all approaches to AI are ML. Not all the constructs studied and applied in the field of AI can learn and change themselves in the process. In fact, not all AI tools are concerned with emulating natural intelligence; some may merely "enable machines to achieve complex tasks in complex environments" (Batarseh, 2018). For example, evolutionary algorithms are inspired by the processes of biological evolution. They can be considered intelligent algorithms because they enable a computer to solve or produce very good answers to complex optimization problems. On the other hand, metaheuristics, such as ant colony optimization, emulates an ant colony's collective intelligence to solve complex problems.

Nevertheless, most optimization and search algorithms do not change their structure in response to the data, at least not in their classical implementations. Instead, they iteratively react to the data in a way that emulates their object or process of inspiration, but the parameters of the algorithm remain constant in a single use. They do not get better at solving problems the more data they have; in fact, they do better with the smaller datasets or instances of less complex problems, even as they were developed to help with problems intractable by traditional analytical methods.

Other methods in the IA toolset go as far as to be defined by their unchangeability, such as expert systems. Expert systems are computer-based systems that can give users suggestions or explanations comparable to an expert. They have three constituent parts: a user interface, an inference engine, and a body of stored knowledge (Holsapple & Whinston, 2013). Their classical implementation assumes that any change to their carefully and systematically collected and processed data originating from human experts would be detrimental to its task. By design, expert systems are supposed to have within them all the knowledge needed to take the user's input or query from their user interface and process it in the inference engine to output a suggestion or explanation comparable to that of the experts who provided the information for the system. Additions to the knowledge stored within the expert system or to its rules in the inference engine are usually the prerogative of its developers.

Other planning, scheduling, optimization, and constraint satisfaction methods can be considered intelligent tools and, therefore, part of AI, but not ML. As we have seen, many tools and methods that are intelligent but not ML are part of the field of OR, defined as "the application of scientific and especially mathematical methods to the study and analysis of problems involving complex systems […] as firm management, economic planning, and the waging of war" (Sivazlian, 2009). However, in recent years even the fields of OR and ML have been in considerable overlap. Now ML approaches have been successfully used in conjunction with heuristics and algorithms in other to produce hyper-heuristics, which are heuristics that use ML to find suitable parameters for the next search iteration or approach or in choosing which heuristic approach to use at each step of a search for an answer to a problem (Cowling et al., 2013). Even expert systems, prominently defined as containing expert knowledge, have been augmented with learning capabilities in recent years (Holsapple & Whinston, 2013), as we would expect a human expert to be open to further learning and broadening his expertise. A Veen diagram of the ML, AI, and OR fields can be seen in **Figure 7**.

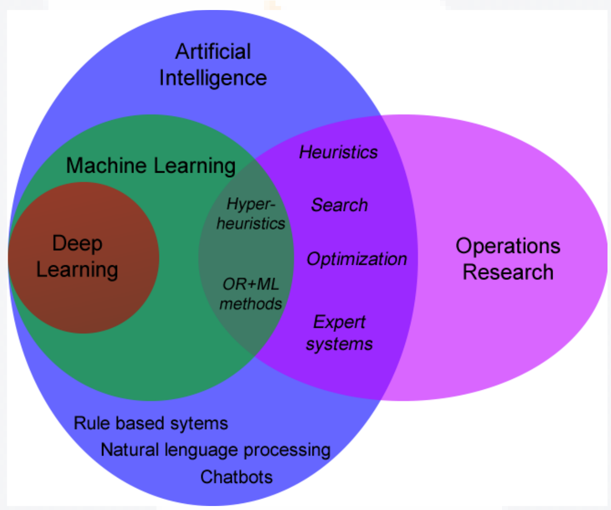


Figure 7: *The fields of AI, OR, and ML showing deep learning as a subfield of ML*

While both the ML and the non-ML approaches can converge in surprising ways in real-world applications, and it could seem that ML is slowly encroaching on all areas of AI, it is important to note that not all problems need an ML approach. Many can be more efficiently solved or approached by the highly developed and researched non-ML approaches of OR and other tools of AI for which a learning ability would either be overkill or unnecessary, as is the case when what is being sought is an optimal solution to a problem and not a prediction or estimate. Also, despite the inspiration from animal and human intelligence from which ML was envisioned at the beginning of its development, in practice it no longer tries to emulate o simulate such intelligences, and the AI field that has kept this paradigm is referred to with a term of its own: Good old-fashioned AI (GOFAI).

GOFAI as a term was devised in 1995 to refer to the non-ML approaches to IA. The GOFAI definition includes things that are AI but not ML. GOFAI "employs programmed instructions operating on formal symbolic representations" (Boden, 2014), hence GOFAI programs are structured by heuristic search and planning, though sometimes using a brute force approach to problem-solving when practical. Further developments were the rule-based systems, of which expert systems are a good example. GOFAI approaches are considered part of AI because their notions of search, planning, and heuristics often do simulate the intelligence of humans. When they rely on knowledge, it comes from specialists in the subjects, as in expert systems, or from users of the languages to be analyzed.

As Boden (2014) explains, the strengths of GOFAI are its ability to model hierarchy and sequential order, allowing for precision in problem-solving and representing specific propositional contents. Its main weakness is its necessary explicitness in that they must be explicitly built to solve specific kinds of problems, and even if that problem were to learn something, the defining features of that something would have to be provided in advance, something that is, most often than not, either impractical or impossible (Boden, 2014). Hence, the fall out of vogue of the GOFAI paradigm.

IV.1.3. Justification of the use of ML

Having seen various methods and options in the AI toolset, it is relevant to see what reasons we may have to use an ML approach to solve a problem. Nilsson (1997b) provides a list of these:

- ***Some tasks cannot be defined well except by example;*** *we might be able to specify input/output pairs but not a concise relationship between inputs and desired outputs. We would like machines to be able to adjust their internal structure to produce correct outputs for a large number of sample inputs and thus suitably constrain their input/output function to approximate the relationship implicit in the examples.*
- ***Important relationships and correlations may be hidden among large data piles.*** *Machine learning methods can often be used to extract these relationships (data mining). Human designers often produce machines that do not work as well as desired in the environments in which they are used. In fact, specific characteristics of the working environment might not be wholly known at design time. Machine learning methods can be used to improve existing machine designs.*

- ***The amount of knowledge available about specific tasks might be too large for explicit encoding by humans.*** *Machines that learn this knowledge gradually might be able to capture more of it than humans would want to write down.*
- ***Environments change over time.*** *Machines that can adapt to a changing environment would reduce the need for constant redesign. New knowledge about tasks is constantly being discovered by humans.*
- ***Vocabulary changes.*** *There is a constant stream of new events in the world. Continuing redesign of AI systems to conform to new knowledge is impractical, but machine learning methods might be able to track much of it.* (Nilsson, 1997b)

These conditions are met by the challenges presented in this dissertation concerning human wellbeing in a social context and the tools of intelligent time management that need to be adjusted to such ever-evolving contexts. As we have seen in previous sections, clear-cut relationships between human actions and wellbeing are hard to come by; instead, it is more promising to analyze large quantities of data for patterns and correlations. Also, the knowledge involving every task done by people covers vast amounts of social sciences and professional knowledge, not only about the practices themselves but about the environments in which they are done, which also change continuously because of technological advances or logistical or even practical concerns. This has the approaches of GOFAI and OR at a disadvantage, except in very specific applications. Finally, this ever-changing nature also applies to the vocabulary used in such environments.

In conclusion, the study of human wellbeing fully justifies using an ML approach, even if aided by some non-ML tools because of the complexity, size, data richness, and changing nature of the problems, not because of some similarity between the intelligence of the ML tools and the intelligence of the people whose data is being analyzed. In fact, it is not a given that the way the ML algorithms solve the problems presented to them will be comprehensible to a decision maker or user, even if the answers are correct, but as we will see in later sections of this dissertation that comprehensibility is certainly desirable if available.

IV.1.4. Common machine learning algorithms

As shown in **Figure 6**, ML methods can be classified by the representation of candidate programs, search algorithm (optimization), and learning method or ML paradigm. Most ML methods use functions to represent the candidate programs, with the prominent exception of decision trees methods and other approaches that use constructs built with general-purpose languages. Many ML methods can use different search algorithms or objective functions without much alteration to their structure. For example, decision trees can use different criteriums such as the Ginni index or Cross-entropy to find the best candidate decision sub-tree to add to the current one. On the other hand, much diversity is had in different ML paradigms or learning methods. A table of different categories of ML models can be seen in **Table 9**.

***Table 9**: Different categories of ML models (Althbiti & Ma, 2020)*

ML paradigm	Task	Type of algorithm	Model
Supervised	Prediction	Regression and classification	Linear regression
			Ridge regression
			Least absolute shrinkage and selection operator (LASSO)
			K-nearest neighbors for regression
			K-nearest neighbors for classification
			(logistic regression)
			One vs. rest linear model for multi-label classification
			Decision trees (DSs)
			Bayesian classifiers
			Support vector machines (SVMs)
			Artificial neural networks (ANNs)
Unsupervised	Features extraction	Dimensionality reduction	Principal component analysis (PCA)
			Singular value decompensation (SVD)
	Description	Clustering	*k*-means
			Density-based spatial clustering of application with noise (DBSCAN)
			Message passing
			Hierarchical

		Association rule mining	A priori

IV.1.5. Training, underfitting, and overfitting

A key concern in ML is how the models fit their training datasets. Most ML models use training datasets to encounter the experience needed to modify themselves into a model that explains the interactions in the data. Suppose the model is fitted correctly for the data and performs well. In that case, it can use its training with the data it has encountered before to make accurate predictions with data not in its original training set. On the other hand, if the training is not done correctly according to the structure and definition of the ML model used, two common undesirable situations can be rise:

- **Overfitting,** which means that the model is so closely fitted to the training dataset and its peculiarities that it is unable to use its training to predict data points not in the training set effectively.
- **Underfitting** means that its training is not sufficient to predict from data points not in the training dataset effectively.

Overfitting is usually solved by trying training with random subsets of the whole training dataset. However, other solutions, such as introducing uncertainties in the model, can work with some ML models. Underfitting is usually solved by feeding the ML model a more extensive training dataset. However, depending on the model, other actions can be taken to overcome this issue depending on the goals of using a particular ML model.

IV.2. Artificial Neural Networks

Artificial Neural Networks (ANN) are an attempt to emulate, to a certain point, not only the intelligence of humans and animals but also the brain's structure. However, much has been learned about the brain since the possibility of modeling neural networks was proposed (Kleene, 1956), and so even as it was predicted at that time, we now know more about how the human brain is a much more complex structure than even the most advanced current ANNs models. However, this divergence from its inspiration has yet to impact their effectiveness or discourage their use as they remain one of the most used and high performing ML models and, when computational resources became more abundant and cheaper, ushered in the era of deep learning (DL). Hence, deep ANNs were the first promising algorithm to test for the purposes of this dissertation.

IV.2.1. Definition of artificial neural networks

An ANN can be defined as a "computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase" (Sammut & Webb, 2017a). As we can see from the previous definition, not all ANN models are adaptive systems that change their structure as they learn, and such models would not be examples of ML approaches. Some small ANNs can be used without training with a random initialization or hardcoded parameters highly optimized for a particular task. Other bigger ANNs, such as weight agnostic ANNs (Gaier & Ha, 2019), can be used without learning if they have already been structurally designed to handle the task.

A more mathematical approach to the definition of ANNs is as "networks of non-linear elements, interconnected through adjustable weights [that] play a prominent role in machine learning. They are called neural networks because the non-linear elements have as their inputs a weighted sum of the outputs of other elements—much like networks of biological neurons do" (Nilsson, 1997a). This mathematical approach is important because it has been demonstrated that ANNs can approximate any function via the composition of such non-linear elements (Goodfellow et al., 2016), which is what, in fact, happens with the function that ANNs are used to optimize, as we will see in later sections. However, that it is demonstrably possible does not mean it is trivial or even easy in most cases. Finding ways to approximate the function better is part of the challenge of using and developing ANN models.

A different view of the characterization of ANNs is as "higher-level abstractions of the classical models that are commonly used in machine learning" (Aggarwal, 2018). This view is important because ANNs have a double heritage, one from their inspiration in the biological neural networks' connectionist approach and another from their inspiration in simpler ML models that take the place of the nodes in a network or graph. According to Aggarwal (2018):

> *A neural network can be viewed as a computational graph of elementary units in which greater power is gained by connecting them in particular ways. When a neural network is*

> *used in its most basic form, without hooking together multiple units, the learning algorithms often reduce to classical machine learning models […]. The real power of a neural model over classical methods is unleashed when these elementary computational units are combined, and the weights of the elementary models are trained using their dependencies on one another. By combining multiple units, one is increasing the power of the model to learn more complicated functions of the data that are inherent in the elementary models of basic machine learning. The way in which these units are combined also plays a role in the power of the architecture and requires some understanding and insight from the analyst.* (Aggarwal, 2018)

The power of this higher-level abstraction is realized when a very powerful and intricate ANN-based ML model, capable of the intricate calculations needed to model complex scenarios, is implemented by curating a very small set of hyperparameters and providing a sufficiently big dataset. ANN-based ML models are infamous for being "black boxes" given the complexity of their internal computations, even as they are made of very simple networked computational nodes. However, one of the most important strengths is the relative simplicity of their use and configuration in modern implementation in a growing variety of general programming languages, such as Python, R, Java, SAS, and others.

IV.2.2. The perceptron, a single artificial neuron model

To better understand ANNs built with several layers of artificial neurons, it is convenient first to understand a model with just one neuron: the perceptron. Which is a binary classification algorithm that makes its predictions with a linear function, as will be made evident in its mathematical expression, and therefore can only classify linearly separable data points such as the one shown in **Figure 8**.

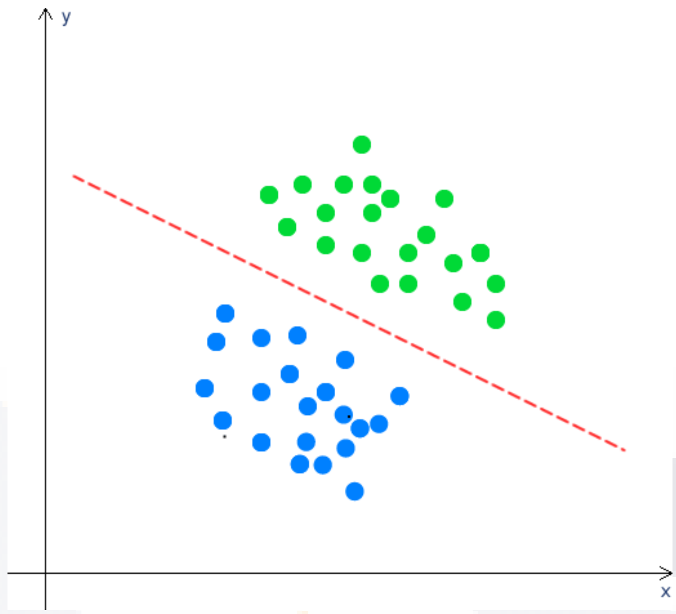


Figure 8: *Linearly separable data and a candidate linear function to divide them*

Historically, the first model of an artificial neuron was the perceptron, also sometimes known as a single-layer perceptron, to distinguish it from ANNs built of several layers of perceptrons. While not all ANNs are built from artificial neurons like the one used in the perceptron model, such ANNs are very common and a starting base for ANNs that use different models of artificial neurons. Originally, the perceptron was not designed as an algorithm but as an electronic machine. Even the objective function of the original electronic perceptron was unknown and had to be reverse-engineered later (Nilsson, 1997a). A photo of the Mark I perceptron computer developed in the 1950s can be seen in **Figure 9**.

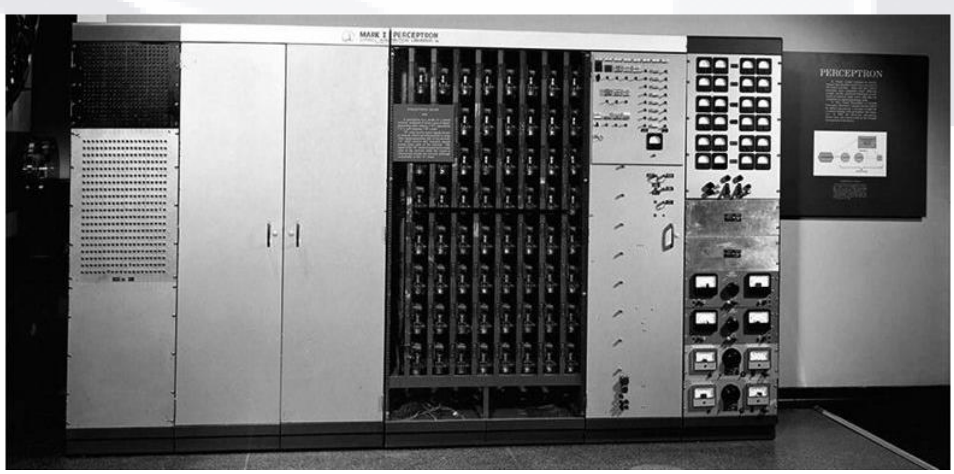


Figure 9: *The Mark I perceptron* (Nilsson, 1997a)

However, even in these times, the perceptron was already fashioned after a simple model of a biological neuron, as seen in **Figure 10**. We can see diagrams of the perceptron in **Figure 11**, and we can see how the dendrites are analogous to the input nodes of the perceptron. The axon is like the output node, which in the case of a multilayered model can have connections to other neurons, as the synapses of biological neurons connect to the dendrites of other neurons, and in the case of single layer models, just outputs a result.

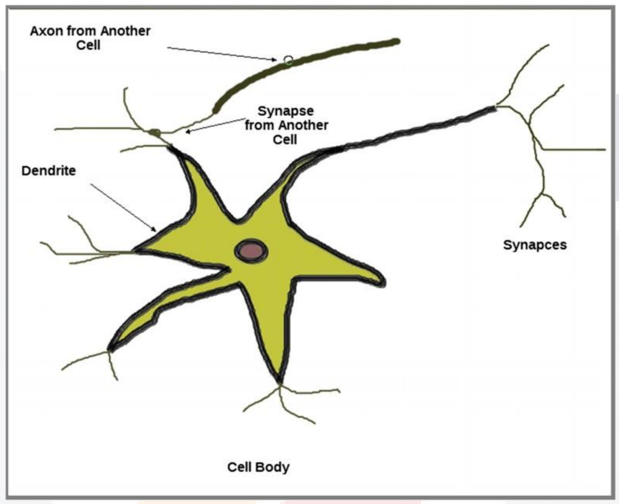


Figure 10: *A simplified diagram of a biological neuron* (Livshin, 2019)

The bias seen in **Figure 11** (b), is simply a predefined invariant value feed to the perceptron as if it came from another neuron before the one in the perceptron model, this can help with some applications such as imbalanced training datasets, such as cases in which all (x, y) points of the dataset exist in a single quadrant, or are very far from the origin $(0,0)$, among other cases.

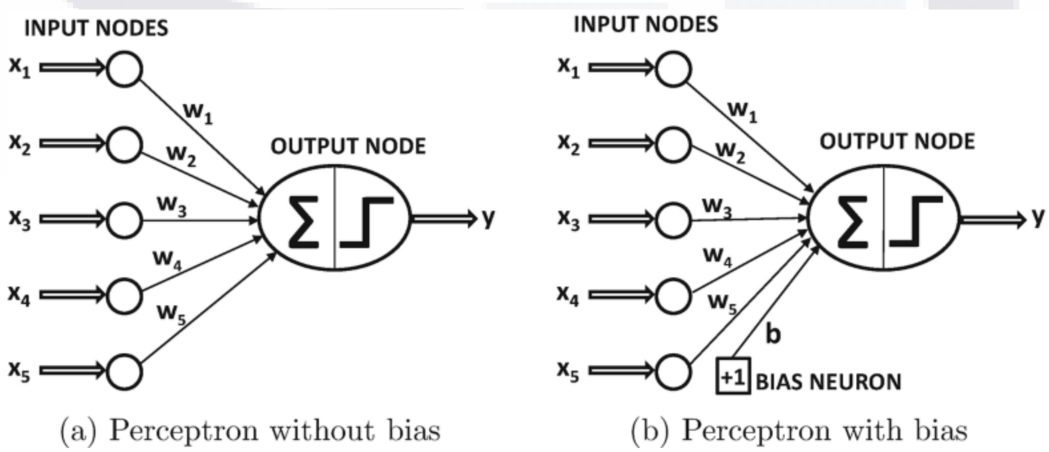


Figure 11: *Diagrams of the basic architecture of the perceptron, without (a) and with bias (b)* (Nilsson, 1997a)

As we can see in **Figure 11**, the input layers contain d nodes that transmit the features of a vector $X = [x_i, \dots, x_2]$, and then, at the output node, each feature is multiplied with a weight from the vector $W = [w_1, \dots, w_2]$ and then a sum is done. Therefore, up until the first half of the output node, the calculation done is the dot product of X and W as shown in equation (3).

$$\overrightarrow{W} \cdot \vec{X} = \sum_{i=1}^{d} w_i x_i \tag{3}$$

The value calculated in in equation (3) is called the pre-activation value. This value is important because sometimes the analyst or researcher will want to compare different implementations of artificial neurons with different post-activation values. The post-activation value is calculated using the pre-activation value and processing with an activation function, of which there is a rich variety.

In the case of a perceptron, a Boolean function is used as an activation function to take the pre-activation value and compute it into either a 1 or 0. While most artificial neurons use such functions, sometimes non-Boolean functions can be used too; for example, in a more complex ANN which needed to predict which digit is depicted in an image, it would use a function with an output range like $\{0,1,2, \dots ,9\}$. However, in the case of the perceptron in its original inception, it is only helpful for a binary prediction or classification, and a Boolean function is used. A common such Boolean function is the sign function, which maps a real value, such as the result from $\overrightarrow{W} \cdot \vec{X}$ into either 1 or -1. Therefore, the post-activation value and output $\hat{y}$ of the perceptron is calculated as shown in equation (4).

$$\hat{y} = sign\left(\sum_{i=1}^{d} w_i x_i\right), \hat{y} \in \{-1,1\} \tag{4}$$

In the case of the perceptron with bias, equation (4) would be modified slightly to include the b constant value we saw in **Figure 11**, as shown in equation (5):

$$\hat{y} = sign\left(\sum_{i=1}^{d} w_i x_i + b\right), \hat{y} \in \{-1,1\} \tag{5}$$

The sign function, in this case, takes the role of an activation function which, as we have seen, is the function that defines the output of an artificial neuron. As we have already mentioned, the sign function is not the only possible activation function; we will see others in the next section of this chapter.

Returning to what we saw in the previous section of this chapter, we know that ML models have three parts: a representation of the knowledge, a search algorithm for candidate programs, and a way to evaluate such programs. In the case of the perceptron, and ANNs in general, the representation of the knowledge are the neurons themselves with its input's nodes, which contain the parameters to be modified: the weights in the $\overrightarrow{W}$ vectors. The change of these parameters in the $\overrightarrow{W}$ vectors is how the learning process modifies the structure of the model.

How the programs are evaluated in the case of ANNs is with an objective function, which in the case of a perceptron, and neural networks, generally, is to be minimized, and it is called a loss function. For the perceptron, the function is as shown in equation (6).

$$\min_{W} L = \sum_{(\bar{X},y)} (y - \hat{y})^2 = \sum_{(\bar{X},y)\in D} (y - sign\{\bar{W} \cdot \bar{X}\})^2 \qquad (6)$$

In which *D* is the dataset containing feature-label pairs. For example, the years of birth of a group of people and whether they hold a college degree, So D would look like $D = [(1955,0), (1991,1), (2001,0), \dots]$. It is to be noted that in the case of this loss function, it is easy to see that it is non-differentiable, given that the sign function is non-differentiable. In practice, such loss functions are substituted for other *smooth* functions that are differentiable because the search algorithm of multiple-layer ANNs uses the gradient of the loss function to look for the best set of w_i values. However, in the case of the perceptron, we can use a simpler algorithm to train it.

A way in which the perceptron learns or finds the “right version” of itself is the following:

1. A training set is had in the format:
    a. $D = \{(x_1, y_1), \dots, (x_s, y_s)\}$ in which each x_j is an n-dimensional input vector, and each y_j a scalar target value.
2. A learning rate r is defined, with $r \in [0,1]$
3. A counter is initialized at $t = 0$.

4. The components of the W vector are initialized to 0 or a small random value therefore $W(t) = [0, \dots, 0]$
5. For each (x_j, y_j) pair in the D dataset, the following is performed:
   a. The $\hat{y}(t)$ output is calculated.
   b. The weights are updated:
      i. $w_i(t+1) = w_i(t) + r\left(y_j - \hat{y}_j(t)\right)x_{ji} \ \forall \ 0 < i < n$
   c. The loss function is evaluated.
      i. If minimized or its value is below a predefined threshold, the process can stop, and the perceptron is considered trained.
      ii. If not minimized or its value is above a predefined threshold, the process goes back to the beginning of step 5. (Nilsson, 1997a)

Once the learning process is done, the perceptron can be considered trained, and an input vector x can be introduced through the input nodes, and a prediction about its y value can be had. It should be noted that this learning process is for the perceptron model only, and ANNs models with more than one neuron use other learning algorithms, such as backpropagation. Also, the evaluation of the loss function can be omitted if the dataset is small and, therefore, all its points are used to train the algorithm. However, in specific datasets, this can lead to overfitting. On the other hand, the model is considered underfitted if the loss function is not minimized or below a reasonable predefined threshold.

IV.2.3. Multilayer and deep learning artificial neural networks

While the perceptron is a one-layer artificial neuron model, nowadays, most ANNs used are multilayer, that is, they use multiple artificial neurons in a multilayer structure. Moreover, in the case of multilayer ANNs, we must clarify that the models we will see in this section, and the ones implemented for the experimental part of the dissertation, are all deep feedforward ANNs, which means that they are multilayered ANNs that do not form cycles. We can see diagrams of a multilayer ANN in **Figure 12** with and without bias and in different notations. Notice that each layer of artificial neurons has its own bias fed into them, which can be the same or different for each layer depending on the nature of the application.

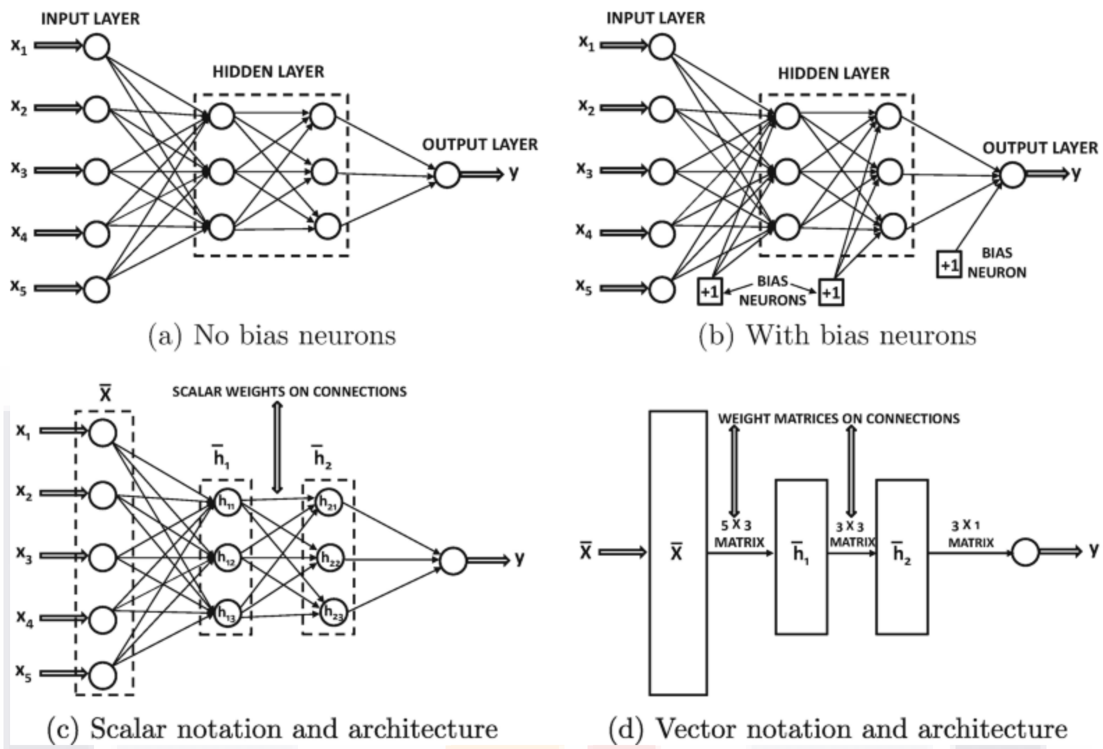


Figure 12: *Basic architecture of a feedforward network with two hidden layers and a single output layer* (Nilsson, 1997a)

We can see that we still have the input layer for multiple features x_i, but now each feature is fed to various artificial neurons in a first layer after the input layer, which now is not the output layer. In this kind of ANNs, from the viewpoint of each artificial neuron they function similarly to the perceptron: receiving inputs, processing them with their own set of weights, making a sum, processing this sum with their activation function, and then sending an output; with the main difference being that this output is sent equally to all the artificial neurons in the next layer. The layers between the input and output layers are called hidden layers.

Multilayer ANNs can have any number of hidden layers, and these hidden layers can have the same number of artificial neurons as each other, or the numbers can differ. These characteristics can obey any number of reasons, some concerning the nature of the data fed to the ANN and others the purpose of the multilayer ANN particular implementation. Also, the output layer is not necessarily limited to just one artificial neuron and, therefore, just one output. An example of an ANN with multiple artificial neurons in its output layer and different numbers of artificial neurons in its hidden layers is the autoencoder, whose structure is a function of both the kind of that data fed to it and the purpose it has which is to “find some latent representation of the points in a training dataset that preserves the information contained in the data points, while simplifying the data in a certain way" (Lempitsky, 2019) and is therefore highly used in computer vision applications; an autoencoder has the same number of input nodes as output ones. We can see a diagram of one in **Figure 13**.

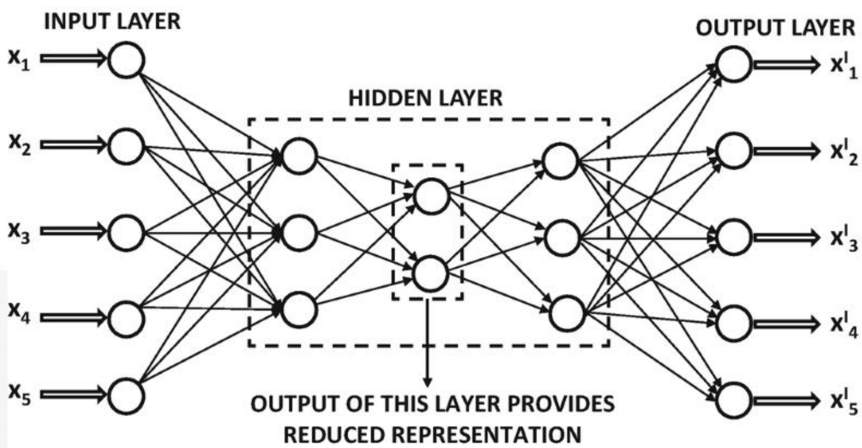


Figure 13: *A diagram of a simple autoencoder* (Nilsson, 1997a)

The process of training a multilayer ANN is far more complicated than with the perceptron because the loss function of the whole network is a "complicated composition function of the weights in earlier layers" (Nilsson, 1997a), and its loss score and gradient needs to be calculated during the process as we can see in **Figure 14**.

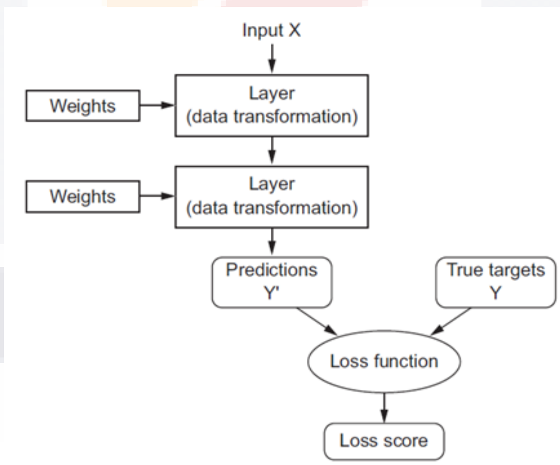


Figure 14: *A loss function measures the quality of the network's output* (Chollet, 2018)

The training of a multilayer ANN is done with the backpropagation algorithm, "an iterative procedure that adjusts network weight parameters according to the gradient of an error measure. The procedure is implemented by computing an error value for each output unit and by backpropagating the error values through the network. [Backpropagation] is a direct application of the *gradient descent* approach to optimization" (Munro, 2010). It is a dynamic programming application requiring two phases: a forward phase in which the errors on the ANN are calculated and accumulated and a backward one in which the weights, or parameters, are updated. This updating of the weight or "adjustment is the job of the optimizer, which implements [backpropagation]. Initially, the weights of the network are assigned random values, so the network merely implements a series of random transformations. Naturally, its output is far from what it should ideally be, and the loss score is accordingly very high. But with every example the network processes, the weights are adjusted slightly in the correct direction, and the loss score decreases. This is the training loop, which, repeated a sufficient number of times (typically tens of iterations over thousands of examples), yields weight values that minimize the loss function. A network with a minimal loss is one for which the outputs are as close as they can be to the targets: a trained network. Once again, it is a simple mechanism that, once scaled, ends up looking like magic" (Chollet, 2018). The feedback loop can be seen in the backpropagation process diagram in **Figure 15**.

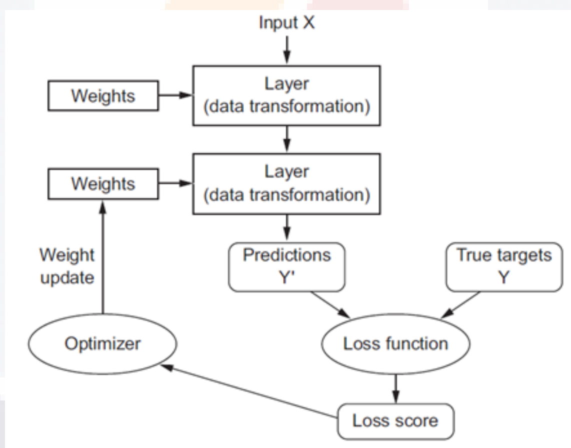


Figure 15: *In the backpropagation algorithm, the loss score is used as a feedback signal to adjust the weights* (Chollet, 2018)

Deriving backpropagation and fully explaining it is an arduous process outside the scope of this chapter, whose purpose is to introduce fundamental concepts of ANNs to understand DL better. However, while most implementations of ANNs and DL have highly efficient and well-designed backpropagation algorithms already coded into them, it should be noted that, apart from overfitting, most of the challenges in adjusting neural networks come from the process of backpropagation: problems with exploding or vanishing gradients, convergence difficulties, local/spurious optima, and other computational challenges (Nilsson, 1997a). Given that ML is a practical approach, a process of trial and error may be taken with new datasets. If problems arise, the nature of the data used to train the ANN must be considered, as well as the hyperparameters of the model, which need to be set before the backpropagation process begins.

IV.2.4. ANN's parameters, hyperparameters, and loss functions

In the context of ANNs, like those composed of multiple perceptron-like artificial neurons, the parameters are said to be the weights encoded in the input nodes of each artificial neuron. It could be said that the goal of training ANNs is to solely to find the right values for these parameters, as we can see in **Figure 16**. Nonetheless, this does not mean that more sophisticated ANNs models may not have other parameters encoded in their artificial neurons or structures. An ANN's parameters are where the model's structure is changed with the learning process.

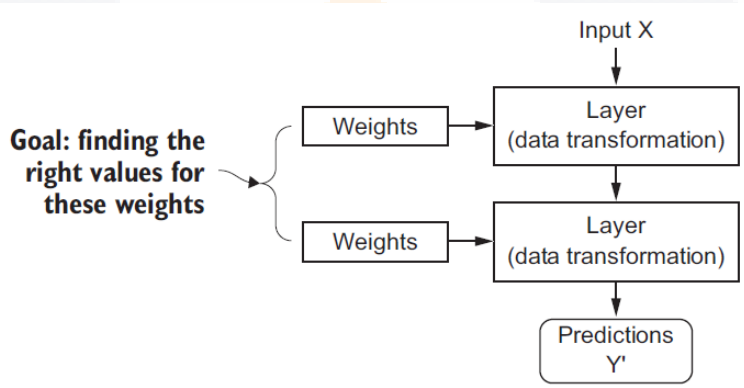


Figure 16: *A neural network is parameterized by its weights and the goal of its training is to find the right values for these weights* (Chollet, 2018)

On the other hand, hyperparameters are values set beforehand “that help in approximating complex relation[s] between input and output” (Ayyadevara, 2018). The hyperparameters are primarily about the structure of the ANN, as we have already seen that we can have different ones in the examples in **Figure 12** and **Figure 13**. Hyperparameters may include:

- Number of hidden layers
- Number of hidden units in each hidden layer
- Number of units in the input layer
- Number of units in the output layer
- Activation function

- Number of epochs
- Learning rate

Notice that the loss function is not a hyperparameter; this is because the loss function is implicit in the architecture of the ANN. Sometimes the loss function is straightforward as with the perceptron, but even small multilayer ANNs can have very complicated and challenging to express loss functions; in fact, for some big ANNs, the loss function may not even exist as an explicit expression, which contributes to the reputation of ANNs as black boxes. There are no set rules or recipes to adjust each hyperparameter, but mere conventions and rules of thumb. Sometimes optimization algorithms such as evolutionary algorithms are used to estimate a set of hyperparameters better than a set chosen by an educated guess. However, the nature of the data and the intention of the implementation of the algorithm must always be considered when choosing hyperparameters.

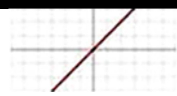
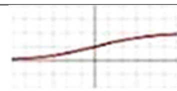
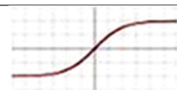
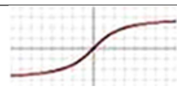
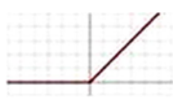
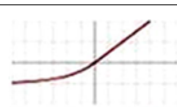
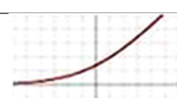
A higher number of hidden layers can enhance the model's representation of the training dataset. However, too many hidden layers can mean more computational time, and there are data-dependent thresholds to how many hidden layers can be added while gaining performance. The law of diminishing returns applies in the setting of this hyperparameter in, sometimes, complex ways.

Equally dependent on the applications of the model and its intention is the number of units in each hidden layer. In most prediction and classification ANN models an equal number of units in each hidden layer can be adequate, however special applications such as the autoencoder seen in **Figure 13**, can need different numbers of artificial neurons in different layers.

The number of units in the input layer mostly depends on the number of features from the training dataset we wish to feed to the ANN. Similarly, the number of units in the output layer is linked to the kind of result that we want; a single value will suffice for a single classification, but multiple values would be needed for applications such as image, video, and audio processing, or models that classify or predict multiple elements at a time.

Also, the choice of activation function "depends on various factors, such as the interval where they are well-behaved (not saturated), how fast the function changes when its argument changes, and simply [personal] preferences" (Livshin, 2019). A set of the most used activation functions can be seen in **Table 10**.

Table 10: *Most used activation functions* (Livshin, 2019)

Name of the fucntion	**Plot**	**Equation**	**Derivative**
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0, & x \neq 0 \\ \nexists, & x = 0 \end{cases}$
Logistic		$f(x) = \dfrac{1}{1 + e^{-x}}$	$f^{\prime(x)} = f(x)[1 - f(x)]$
TanH		$f(x) = \tanh(x)$	$f'(x) = 1 - [f(x)]^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \dfrac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$
Parametric Rectified Linear Unit (PReLU)		$f(x) = \begin{cases} \alpha x, & x < 0 \\ x, & x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha, & x < 0 \\ 1, & x \geq 0 \end{cases}$
Exponential Linear Unit		$f(x) = \begin{cases} \alpha(e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha, & x < 0 \\ x, & x \geq 0 \end{cases}$
Soft Plus		$f(x) = \ln(1 + e^x)$	$f'(x) = \dfrac{1}{1 + e^{-x}}$

As a general recommendation, "using a specific activation function depends on the function being approximated and on many other conditions. Many recent publications suggest using $tanh$ as the activation function for hidden layers, using the linear activation function for regression, and using the $softmax$ function for classification" (Livshin, 2019); however, as with the other hyperparameters, experimentation and consideration of the nature of the data and the intention of the implementation of the ANN model is a good approach if the performance of the model is critical, and if not, the previous general recommendations will suffice. The number of epochs or backpropagation iterations that the ANN will perform to train is also a hyperparameter that can significantly enhance the model's performance. A few tens of epochs might be enough for a first attempt, but more epochs may be needed if the data is home to complicated patterns.

As for the learning rate, this hyperparameter “controls how much we are adjusting the weights of our network with respect to the loss gradient” (Zulkifli, 2018). Again, careful consideration should be taken when choosing this hyperparameter because if it is made “too small, convergence will be very slow, but if we make it too large, the method can fail to converge at all” (Murphy, 2012). There are implementations of ANNs that use different learning rates for different parameters. There are some methods to calculate reasonable learning rates depending on the architectures of the ANN and the training data set. However, for most current implementations of ANNs the default values are usually sufficient for a first trial, and if the ANN is not learning well or is taking much time to train, this hyperparameter is a good place to start adjustments.

IV.3. Deep Learning

Deep learning (DL) is one of the most promising subfields of ML, but what it is and how it differentiates from other ML approaches is sometimes challenging to grasp. Mostly because disinformation about DL comes from two almost contrary poles: on one side, many overhyped promises of fully automated complex tasks (such as driving cars or developing software) taken by artificial minds with a deep understanding of their environments and an equally deep human-like capacity to adapt. On another side, we have a deficient characterization of DL as merely a subfield of ANNs with lots of hidden layers. In this section, we will clarify what DL is, how it is practiced by professionals and researchers, and the most used varieties of DL models and their applications. This clarification is necessary because the ANNs tested in this dissertation are deep feedforward ANNs, and, therefore, part of the domain of DL, which was a very promising approach to the complex interactions found in the dataset analyzed.

IV.3.1. Definition of deep learning

DL can be more specifically defined as “a specific subfield of machine learning: a new take on learning representations from data that emphasizes learning successive layers of increasingly meaningful representations” (Chollet, 2018). Notice that no mention of ANNs is made in this definition, despite the DL field being dominated by them. If we have this expectation, it is because we confuse the overwhelming way professionals and researchers practice in the field of DL with what the DL field is as a whole.

Another similar take is that “deep learning is a class of machine learning techniques that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification” (Akerkar, 2019). This definition is similar to how ANNs are defined as “networks of non-linear elements, interconnected through adjustable weights” (Nilsson, 1997a) as we saw when we covered them in their own section; however these interconnected non-linear elements need not need to be artificial neurons, and it can be argued that since we already know that even artificial neurons are a crude model of biological ones, it would be best to define DL in a more mathematical way.

The deepness of DL is not a simile for deeper thought or even deeper math for that matter. In fact, the deep in DL refers to the “successive layers of increasingly meaningful representations […], it stands for this idea of successive layers of representations” (Chollet, 2018), so the deep in DL refers to the depth of all these successive hidden layers in the model that capture the information learned in increasing meaningfulness, and that is why a model with a single hidden layer is referred to as a shallow model.

Concerning meaningfulness, a warning should be issued: the meaningfulness of information is always in its interpretation. For example, in a properly implemented DL model tasked with creating the image of a human face, the image produced through the successive layers will be more and more recognizable as a human face by the users of the implementation to the point when this interpretation will have a meaning in the minds of these users: this is the face of a human with such and such characteristics. In fact, the same can be said of the solutions produced by other ML models, whether these outputs are a numeric vector or a single yes/no answer. However, this clarification is more relevant in the context of DL.

Chollet points out that other possible names for the field of DL could have been “layered representations learning and hierarchical representations learning” (Chollet, 2018). The perspectives emphasized in these names better allow for the currently rare cases in which these learning representations are not learned via ANNs. However, as these approaches are rare nowadays, the present field of DL is pretty much almost concurrent with ANNs; yet, it should be noted that what defines the field of DL is that it is the approach of successive layers of increasingly meaningful representations and not that these layers are made of artificial neurons, which happened to be a handy building block for DL models because of them standing in the powerful confluence mentioned by Nilsson (1997a) of very simple models connecting to each other in order to form a powerfully capable layered network, and the resulting abstraction of the whole model facilitating its application in practice.

Chollet (2018) offers an example of a trained DL model identifying, in successive layered representations, what digit is an image composed of pixels in **Figure 17**.

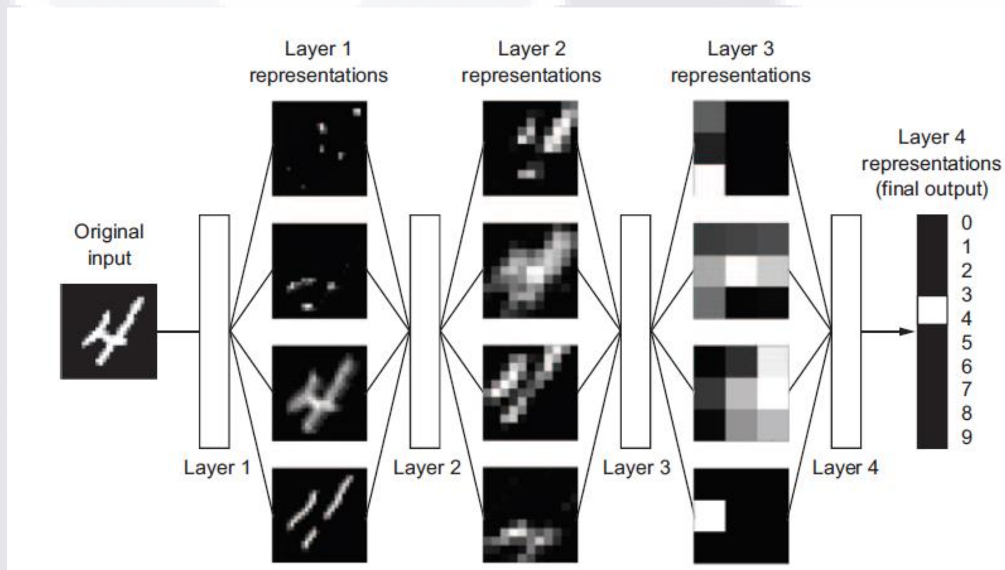


Figure 17: *Example of deep representations learned by a digit-classification model* (Chollet, 2018)

As we can see in the diagram of **Figure 7**, DL is a subfield of ML and not a subfield of ANNs. If we mapped the three fields, they would look more like in **Figure 18** where we can see that there are a few ANNs applications that are not part of DL and rare cases of DL that do not use ANNs. We can also see that we could classify other non-deep ML algorithms that use a single layer as of representation as "shallow learning." Examples of shallow learning would be support vector machines, decision trees, and all the regression-based approaches with a single layer. It is to be noted that ensembles are not the same as layered representations, so an assemble of decision trees or support vector machines or even of perceptrons or shallow ANNs would not be equivalent to a DL approach.

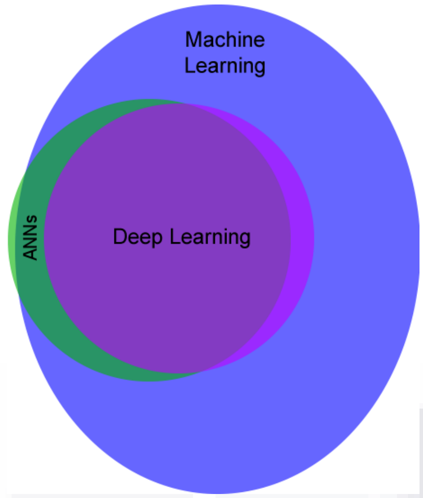


Figure 18: *A Veen diagram of the fields of ML, DL, and ANNs as currently practiced.*

Hence the deep in Deep Learning denotes an entirely particular approach in ML and is not merely ANNs with more hidden layers. In fact, from this perspective, all the other ML models mentioned in this dissertation could be said to be "shallow."

IV.3.2. Practice of deep learning

Having clarified the difference between DL, ML, and ANNs, we should point out that the confusion about these is not unwarranted given that by DL, many, if not most current references do mean a subset of ANNs models. This confusion remains because, as we have mentioned, DL models that do not use ANNs are scarce, as well as implementations of such models in libraries of general-purpose languages or analytics software. Therefore, as this section's objective is to clarify the definition of DL only to shed light on the reasons deep feedforward ANNs were tested in this dissertation we will go no further into non-ANN DL models.

On the other hand, all the material covered in the previous section on ANNs about definitions, structures, loss functions, activation functions, parameters and hyperparameters, and the backpropagation algorithm apply equally to the multilayered feedforward ANNs already covered, and to the other kinds of ANNs in DL we will mention in this section that have architectures that allow for cycles or other more complex structures and patterns of training. The main difference between DL and non-DL ANNs is that the first have more hidden layers to classify as deep. All the different structures in DL models are only possible because of their relatively big number of hidden layers.

There is no definitive inferior limit to the number of layers a DL model must have to qualify as DL. However, practical assumptions and the implementation and use of DL libraries in general-purpose languages assume that a minimum of two hidden layers in an ANN model qualify as DL. This convention, however, may change in the future. On the other hand, an ANN model with just one hidden layer would be universally called a shallow model, and no hidden layers would mean returning to a perceptron model or something similar to it. Therefore, it is helpful to note that there is a tendency in most recent literature and material produced by the professional community that uses ANNs to assign the DL name to practically all but the most simple current applications of ANNs, as we can see that in the diagram of **Figure 18** the non-DL ANNs are a slim sliver of the whole ANNs field that is currently dominated by the DL approach. So, nowadays, for most practical and professional purposes, to talk about DL is to talk about ANNs.

In the same vein, deep ANN models are also characterized by having large numbers of artificial neurons on each hidden layer, but similarly as with the case of the number of hidden layers, the only inferior limit to the number of artificial neurons in a hidden layer is practicality: a hidden layer in a DL model may have two or more artificial neurons.

IV.3.3. Justification for the use of deep learning

The briefer answer as to why use DL models is “because they work.” Nevertheless, a longer answer reveals how the complications and unknowns about ML, in general, become more important as more and more performance is demanded from these models. ML is itself a very engineering-oriented field, ruled by what works in industry applications and not by what can be demonstrated with a rigorist mathematical approach. There exist generalized universal approximation theorems for any number of hidden layers in ANNs (Goodfellow et al., 2016), which in theory, means that even a shallow neural network can approximate any function via composition. Nonetheless, ANNs with more than one hidden layer work better, as shown in **Figure 19**, thought apparently subject to a rule of diminishing returns.

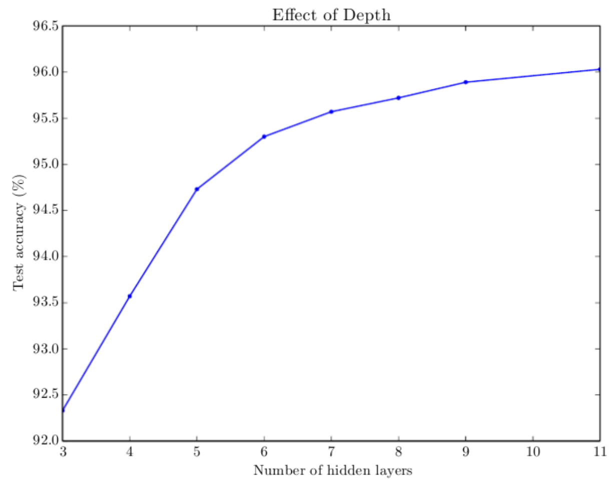


Figure 19: *Results show that deeper models perform better than shallower ones in the task of transcribing multidigit numbers from photographs* (Goodfellow et al., 2016)

Why do deep ANNs work better? Goodfellow et al. (2016) propose three theories: Firstly, theoretically, a shallow model could perform as well as a deep one; however, it would need so many nodes in its lone hidden layer as to be impractical and more computationally intensive than a deep model. Secondly, we do not have suitable training methods to train a shallow model that could perform as well as a deep one. And thirdly, that the architecture of shallow models does not fit the problems that the users of DL are currently interested in.

Interestingly for users of deep models, there has been evidence for quite some time “that shallow nets are capable of learning the same function as deep nets, and in some cases with the same number of parameters as the deep nets” (Ba & Caruana, 2014), however, this is done by training first a deep model and then having the shallow model train by mimicking the deep one. Attempts to train the shallow models directly from the dataset have been unsuccessful. This sequence means that even if we wanted to use shallow models to model complex phenomena, we would still need to begin with deep models to do the heavy lifting of the learning and only then have a shallow model learn from the deep one. However, we just do not currently have the training methods to train a shallow network to high performance from scratch as well as we can with deep networks, or perhaps the problems thrown at DL models are the ones that do not lend themselves to be trained directly by shallow networks.

The issues of why deep models continue to be more useful to model data and relatively easier to train to high performance than shallow ones are far from settled, as are the plateaus in performance as more hidden layers are added to a model in many tasks, as we can see in **Figure 19**; but what we do know is that the depth of DL models, and the different network structures they allow, are currently the only way to model complex phenomena such as image, video, speech, and even human behavior. The explosion of their applications is proof of that. Hence why deep ANNs were the first to be tested for wellbeing research in this dissertation.

IV.3.4. Common types of deep learning

Many relevant and highly used DL models are modifications to the structures we have seen, such as the one in **Figure 12**. The most used and studied are recurrent neural networks (RNN) and their more advanced version, long short-term memory networks (LSTM); as well as gated recurrent units networks (GRU). These were all developed to boost the performance of analysis of sequential data by giving the DL models a kind of memory of what they have processed before. Other very used DL models are convolutional neural networks (CNN), autoencoders, deep belief networks, and deep reinforcement. However, the simplest, but very useful, kind of DL ANN model is the "vanilla" deep feedforward networks like the ones we saw in the previous section, only with two or more hidden layers, as we can see in **Figure 20**, and usually around ten or a few tens.

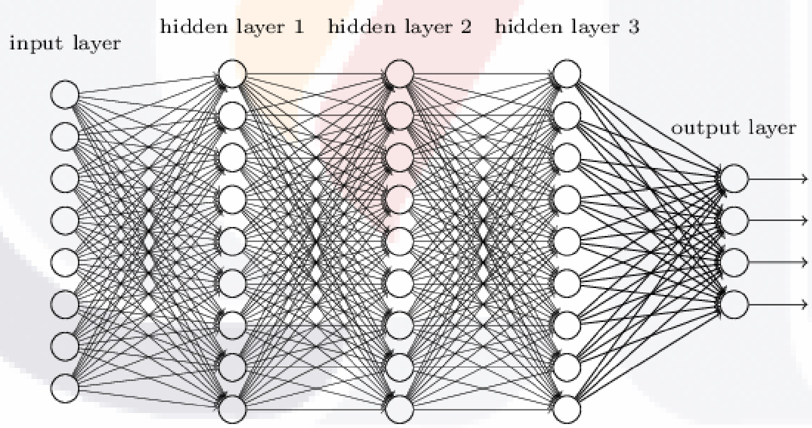


Figure 20: *An example of a feedforward deep neural network with three hidden layers* (Nielsen, 2015)

As we have seen, deep feedforward ANNs can approximate a complex function from their training datasets and are, therefore, good at all the everyday tasks of ML, such as regression, prediction, categorizing, and clustering. They are very good at pattern recognition and, therefore, suitable for applications to read text from images, recognize objects in photos and videos, and particular sounds in audio, leading them to become applications to recognize speech. They are also used to predict how new objects would look in an image that does not contain it, as well as particular sounds to an audio file, which leads to applications in photo and video editing and even artificial speech that can mimic how a human person talks given a script. However, feedforward deep ANNs have problems with data that only makes sense in specific sequences such as speech. That is why these ANNs can recognize words and sentences but cannot construct their own easily hence why architectures like RNNs were developed. As these kinds of instances do not arise in the datasets studied in this dissertation only deep feedforward ANNs were used and tested.

IV.4. Decision trees

Decision trees (DT) employ a different ML paradigm than other highly regarded and used ML algorithms such as support vector machines or ANNs. Instead of using parameters and processing numerical data only, decision trees use a rule-based approach to ML that can process non-numeric data, such as categories, characters, or strings (Joshi, 2020). Because of these characteristics, DT are mostly known as classification algorithms, but they can also be used for regressions. Also, because of their rule-based approach, trained decision trees function in a way that is easily explainable even to a non-technical audience and, therefore, has a relatively high level of interpretability, so much so that they are used as surrogate models to explain other black box models such as DL ones (Blanco-Justicia Alberto and Domingo-Ferrer, 2019). For this dissertation, the interpretability of decision trees-based classification algorithms was a critical advantage that made them one of most attractive and promising algorithms to be tested and used, particularly for the purposes of creating ML applications that are not only auditable but also allow for further mathematical modeling and processing.

This chapter will review the DT's antecedents, definitions, and explanations. We will also see some very used assembles of multiple instances of decision trees.

IV.4.1. Past and present of decision trees

Breiman (1984) and his colleagues presented DT as a methodology using tree graphs, which were not new then but had not been used as a method of classification or regression. In fact, they considered this usage of tree graphs "unthinkable before computers […] unlike many other statistical procedures which were moved from pencil and paper to calculators and then to computers" (Breiman et al., 1984). However, their impact was not significant in Statistics, and it was not until their introduction to the ML literature that they became one of its most used classification methods (Dobra, 2009a). DT were introduced to the ML literature later as an approach to synthesizing decision trees as a knowledge-based system by inductive inference, systems that had been successfully demonstrated in practice (Quinlan, 1986).

Nowadays, decision trees have been eclipsed in use and popularity in the ML community by DL algorithms because the latter have more accurate results in the kinds of tasks more popular in the current practice of ML in the industry. However, the advantages of decision trees' interpretability and relatively easy way of explaining how they work to non-technical audiences have kept them current. Also, they are a high-accuracy option for datasets produced by measurement instruments that collect data on people's decisions, modeling accurately the non-linear and complex interactions found in these datasets (Dobra, 2009b; Joshi, 2020), including those that contain categorical attributes (Origel-Rivas et al., 2020). Given that these are the kinds of datasets that we analyze in this dissertation, decision trees were always a promising option, along with DL algorithms based in ANNs.

IV.4.2. Definition of decision trees

DT are "probably the most intuitive data classification and prediction method. It is also used frequently. While most of the data mining methods […] are parametric, decision tree is a rule-based method […]. A tree is composed of nodes and the leaves are the bottom nodes. At each node except for the leaf nodes, a decision must be made to split the node into at least two branches" (Zhou, 2020). Practically all uses of DT in ML employ a split in only two branches at each node, and so do the DT explained and used in this work. A graphical example of a decision tree can be seen in **Figure 21**.

It should be observed that "while other approaches [to ML] start by writing equations about the properties of data, decision trees start with drawing a tree-type structure such that at each node, there is a decision to be made. At heart, decision trees are heuristic structures that can be built by a sequence of choices or comparisons that are made in a certain order" (Joshi, 2020). This building of the tree is the process known as the training of the tree, which uses recursive partitioning, a process we will see in more detail later in this section.

IV.4.3. The basics of decision trees

A simple example can be made of a problem of classification of books. We could first ask, "Is the book about a fictional matter?" and we would have our first big division between fiction and non-fiction. Nevertheless, we would still have, for example, poetry books along with textbooks on one side because poetry is written about historical facts as well as about fiction and metaphors. From the books already classified as non-fiction, we could ask "Is the matter of the book technical?" to obtain a subset of books that will include textbooks and reference works. From this subset of books, we could continue to ask if the books are about specific objects of study to differentiate subsets of books of sciences from handbooks and other nonscientific material. From the set of fiction books, we could classify by genre by asking about their respective tropes, such as "Does the book include magic?" to classify some books as fantasy and "Does the book involve interstellar travel" to classify others as Science Fiction. A diagram of this decision tree can be seen in **Figure 21**.

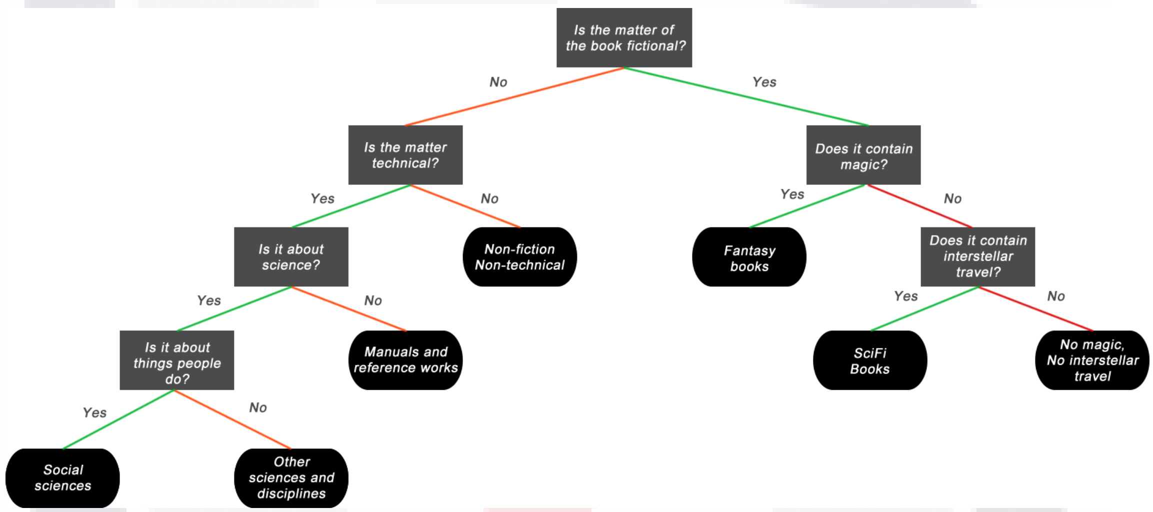


Figure 21: *An example of a decision tree classifying books by their contents*

Joshi (2020) warns us that the actual building process for a decision tree in a generalized setting is much more complex than examples simple enough to explain the basics of decision trees, and this complexity resides, evidently, in finding an efficient order and effective kind of questions we could ask for a decision. The attributes of the training dataset determine the kinds of questions we can ask. In the example of **Figure 21**, it is implied that we could only ask if a book depicts fictional matters or if it contains descriptions of magic if such attributes were contained in the dataset to begin with, for example, a catalog with labels about what the books contain.

One of the most important characteristics of decision trees methods is their interpretability. That is, once a decision tree is trained, it can be graphically explored or audited in order to see how it arrives to the results it achieves. For example, in the example of the DT in **Figure 21**, a set of books can be classified into a few classes, but unlike a classification algorithm based in ANNs a user can "open the box" of the classifier and see how a particular book ends up in a particular class. This characteristic of DT-based models is so powerful and useful that it is used to create explainable models by training DT-based models to mimic DL models in order to explain these last (Blanco-Justicia & Domingo-Ferrer, 2019).

In the book classification example, we could argue that most people familiar with a library or some basic types and genres of books can ask relatively effective questions, even if they would find discovering an effective order for the questions more difficult. However, this is not so with most datasets. Separating effective and non-effective questions and their sequence would require broad domain expertise or extensive knowledge about the structure of the training dataset, which would defeat the purpose of using an ML approach and put us in the realms of GOFAI-like requisites. For example, discerning what kind of questions and in which order to ask them to classify respondents to a time use survey by ethnic group or life satisfaction levels having access only to their activities during a week would require significant knowledge about social sciences, behavioral economics, or about the sample or population studied. This difficulty is why approaches such as recursive partitioning are used to split the nodes of decision trees in an ML training context.

IV.4.4. Recursive partitioning

Despite the similarity of the process of interpreting a trained tree with the thinking process of a person tasked with the same kind of categorizing problem, given the sizeable datasets used to train DTs, they are built using iterative methods that use statistical and logical criteria to determine effective questions to ask at every step. They apply these criteria each time they are about to add a new node to the tree, and recursively try different attributes to ask questions to split nodes. They do this all the way down from the top node or root, which is the first node, into the bottom nodes or leaves, which are the final nodes in which a result has been reached, and no further questions will be asked because all data points that have reached that node have been classified or some stop criterion has been reached. This process is known as the top-down induction of decision trees (TDIDT) or recursive partitioning.

Given today's available computer power, there could be an argument for a random or brute-force approach to partitioning, but Joshi (2020) cautions us that when the data is extensive and high-dimensional, this random or brute-force approach can never be practical.

Since the introduction of decision trees to the ML literature, several efficient criteria for selecting the attribute from which the question will be asked at each step have been developed and implemented (Joshi, 2020). For instance, “how a node is split in a decision tree [can be] based on the entropy computed with the data inside the node. Entropy represents how ‘pure’ the data inside a tree node is. The higher the entropy, the less pure the data is” (Zhou, 2020).

Measuring entropy within a node is important because “if the data are more skewed to a class, the entropy (H) is lower. Decision tree methods tend to pick the attribute that generates the least entropy to split a node. This is easy to understand. Suppose there are two paths before us, and we need to select one to reach our destination. If both paths offer us the same likelihood to arrive on time, we will find ourselves hesitating ($H = 1$[, or maximum entropy]). If one way is much more likely to allow us to arrive on time, our decision won’t be difficult. Moreover, if there is only one way, our selection is definite ($H = 0$)”. (Zhou, 2020)

Zhou (2020) gives an example of a problem using the cross-entropy (H) measure, calculated as shown in equation (7):

$$H = -\sum_{k=1}^{1} P_K \log_2(P_k) \tag{7}$$

In the equation above P_K is the probability for the k^{th} class to occur; it is used to weigh the $log_2(P_k)$. For example, suppose there are 10 data items, among which 3 are “yes” and 7 are “no .”Since there are only two distinct classes (yes or no), m = 2.

For “yes”, p = 3/10 = 0.3, $log_2(p) = −1.74$, $plog_2(p) = −0.52$.

For “no”, p = 7/10 = 0.7, $log_2(p) = −0.51$, $plog_2(p) = −0.36$.

H = −(−0.52 − 0.36) = 0.88.

If there are 10 “yes” and 0 “no,” the computation would be:

For “yes”, p = 10/10 = 1, $log_2(p) = 0$, $plog_2(p) = 0$.

For “no”, p = 0/10 = 0, $plog_2(p) = 0$ (assume $log_2(0)$ is not an error).

H = −(0 + 0) = 0.

If there are 5 "yes" and 5 "no," the computation will give out H = 1.0. (Zhou, 2020)

Let us add to the example above put forward by Zhou (2020). Suppose that in the first case, the one with 3 "yes" and 7 "no" is just one of two attributes in the dataset, the other being two other binary choices between "up" and "down," and "right" and "left." Suppose we have 1 "up" and 9 "downs" and 5 of both "right" and "left." The computations would be as follows:

For the "Yes/No" attribute:

For "yes", $p = 3/10 = 0.3$, $\log_2(p) = -1.74$, $p\log_2(p) = -0.52$.

For "no", $p = 7/10 = 0.7$, $\log_2(p) = -0.51$, $p\log_2(p) = -0.36$.

$H_{yes/no} = -(-0.52 - 0.36) = 0.88$.

For the "Up/Down" attribute:

For "up", $p = 1/10 = 0.1$, $\log_2(p) = -3.32$, $p\log_2(p) = -0.33$

For "down", $p = 9/10 = 0.9$, $\log_2(p) = -0.15$, $p\log_2(p) = -0.16$.

$H_{up/down} = -(-0.33 - 0.16) = 0.49$.

For the "Left/Right" attribute:

For "left", $p = 5/10 = 0.5$, $\log_2(p) = -1$, $p\log_2(p) = -0.5$

For "right", $p = 5/10 = 0.5$, $\log_2(p) = -1$, $p\log_2(p) = -0.5$

$H_{left/right} = -(-0.5 - 0.5) = 1$

Therefore:

$H_{left/right} > H_{yes/no} > H_{up/down}$

We can see that the entropy of the "up/down" attribute at this node is less than that of the "yes/no" and "right/left" attributes, therefore the recursive partitioning algorithm would choose to split this node with the "up/down" attribute. The question asked at this node would be "Is it up?" and not "Is it yes?" or "Is it right?". Intuitively we can see that this is the better option given that with the "up/down" split, we end up with only one element at the "yes" side of the split, which means we arrive at a final node or leave at this early stage. The following nodes are all "downs" that can now be further classified by the "yes/no" attribute, which would still have the least entropy given that the "right/left" attribute would still have maximum entropy (H=1). This would give us a tree as shown in the diagram in **Figure 22**.

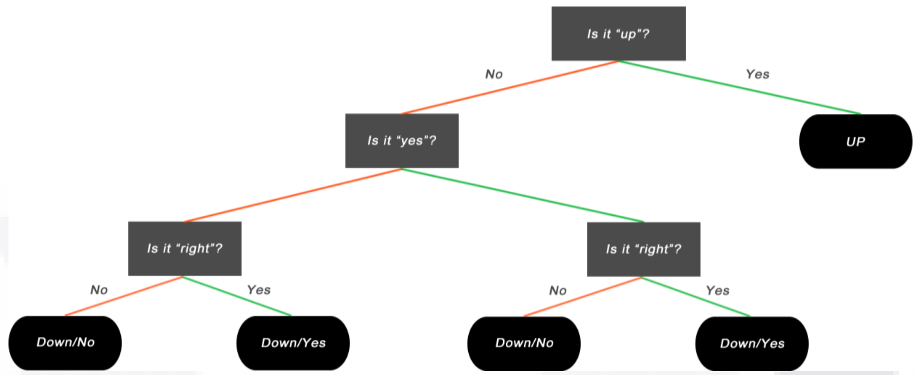


Figure 22: *Tree using an "Up/Dow" and "Yes/No" dataset, doing the split at the root node with the option with the less entropy*

If, on the other hand, we were to go for the maximum entropy split, we would see that we would end up with a bigger tree, which means more computation and less efficiency even if the same result were achieved. At the first split, we would have the same number of elements on each side, not only not achieving a final node at this stage, but both sides of the split still needed to be further processed instead of just one side. This alternative tree is shown in the diagram in **Figure 23**.

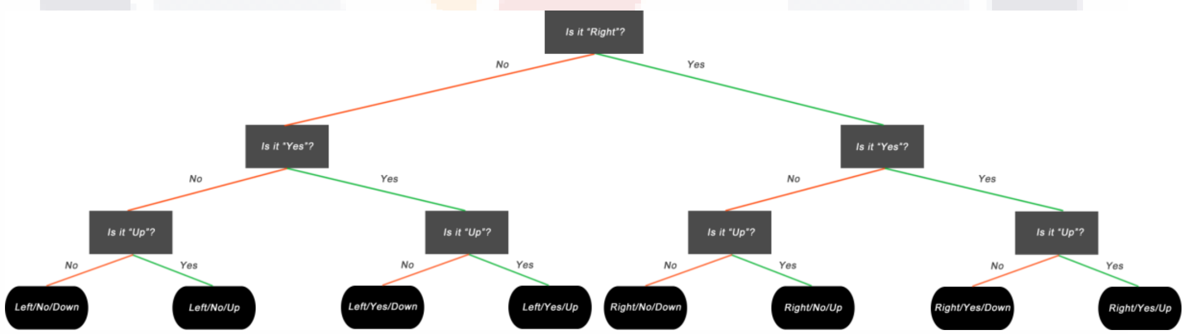


Figure 23: *Tree using an "Up/Dow" and "Yes/No" dataset, doing the split at the root node with the option with the most entropy*

One could question: but in the tree in **Figure 22**, we do not know if the "up" element in the dataset is "up/yes/right," "up/yes/left," "up/no/right" or "up/no/left," but the important thing is that what makes unique that entry in the dataset is that it is the only "up" entry and therefore, using that piece of information is the right way to classify it, and it should be done as soon as possible in the process. Once the results are in, we can see the complete nature of that lone “up” element, but we do not need to know it to classify the element efficiently.

On the other hand, we can see that in the tree in **Figure 23**, three of the leaves with the “up” attribute would not exist or be empty, achieving at that late stage the lone “up” leave that the tree in **Figure 22** achieved at the split in the root node. In a dataset with far more attributes the complexity reduction by choosing the split with the least entropy would be more acute.

Moreover, with the recursive partitioning algorithm, we could use other criteria to calculate the purity, such as the Ginni index, which will be covered later.

IV.4.5. Justification for the use of decision trees

DTs are not currently the most accurate prediction algorithm in several domains and can be very sensitive to small changes in the training datasets (Torgo, 2017). So why use decision trees? Joshi (2020) enumerates some advantages of decision trees algorithms:

1. *More human-like behavior.*
2. *Can work directly on non-numeric data, e.g., categorical.*
3. *Can work directly with missing data. As a result, data cleaning steps can be skipped.*
4. *Trained decision tree has high interpretability compared to the abstract nature of trained models using other algorithms like neural networks, SVM, and other popular ML and DL models.*
5. *Decision tree algorithms scale easily from linear to non-linear data without any change in core logic.*
6. *Decision trees can be used as a non-parametric model. Thus, hyperparameter tuning becomes unnecessary.* (Joshi, 2020)

Also, DTs can be built differently if they are to be used for regression or classification; and different criteriums for selection attributes for new nodes can be used.

The applicability and importance of DTs in practice cannot be overstated. As has already been mentioned, once trained their hierarchical decision-making process is similar to human behavior in real situations, as we saw in the example of **Figure 21**. Therefore, their results are more easily interpretable, even when using multiple DT assembles to boost robustness and performance (Joshi, 2020). In this case, a representative tree for the whole assemble may be selected (Weinberg & Last,

2019) and then interpreted. Also, because of this similarity to human behavior in decision-making, DTs can be more helpful in modeling data obtained from human behaviors and decisions, such as personality profiles and time use data (Marin & Ponce, 2020).

Apart from their relatively superior interpretability compared to other ML methods, Chipman (2011) characterized both their biggest strength and weakness as:

> *The structure of [recursive partitioning] models enables them to identify interactions. For instance, in [**Figure 24**], we see an interaction effect between X_1 and X_2: If $X_2 = \{A, B\}$, then response y decreases with increasing X_1. If $X_2 = \{C, D\}$, then response y increases with increasing X_1. This characteristic is the greatest strength of [recursive partitioning] models, and one of the reasons they are used for exploratory data analysis.*
>
> *This strength is also a weakness. If the relation between predictors and response is additive, very large trees will be needed to capture this relationship. For instance, if $y = x_1 + x_2 + x_3 + x_4 + x_5 + error$ then a tree with 32 terminal nodes will be required to approximate this function with a single step along each of the five predictor axes.* (Chipman, 2011)

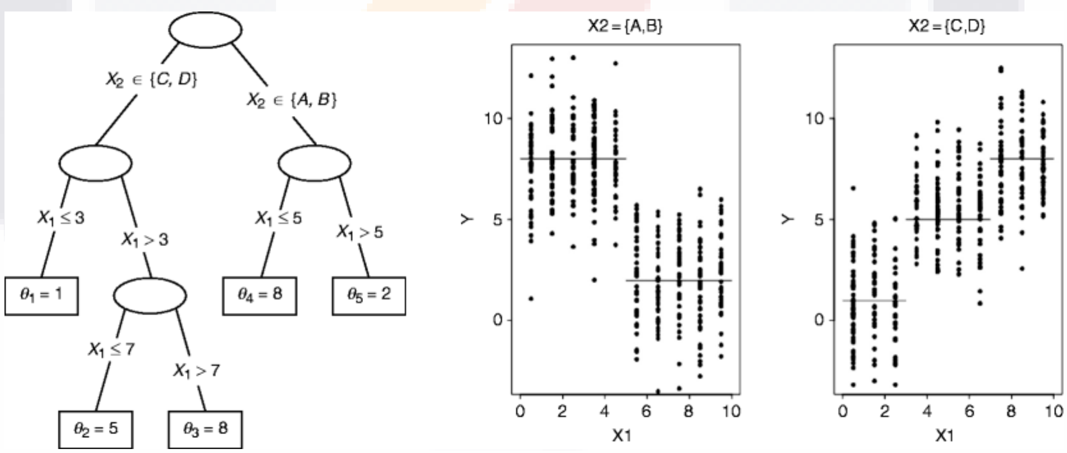


Figure 24: *A regression tree where y ∼ N(ϑ, 22) and x = (x 1, x 2) and a realization of 800 observations sampled from it* (Chipman, 2011)

IV.4.6. Types of decision trees and their construction

As most implementations of decision trees use the TDIDT approach to build the tree, they are usually differentiated by their use (regression or classification) and the method used to iteratively select attributes to ask questions at new nodes. The most common methods are CART, ID3, and CHAID, which can be used for regression and classification. ID3 and CART are very similar and differ in that

CART uses the Ginni index, and ID3 uses cross-entropy as the criterion to split nodes. The example in **Figure 22** would be an example of ID3. On the other hand, CHAID is a different approach that must be studied separately. In the case of regression, the mean square error can also be used as a criterium.

Both CART and ID3 decision trees are similar because when Quinlan (1986) introduced decision trees to ML literature, he presented ID3 as a "variation of CART methodology with slightly different use of optimization method" (Joshi, 2020). On the other hand, the CHAID algorithm is a different approach that uses the chi-square distribution to compare a sample with a population.

As no regression DT techniques are used in this dissertation, we will focus on classification trees. A DT can be characterized as "a tree-structured classification model, which is easy to understand, even by non-expert users, and can be efficiently induced from data" (Fürnkranz, 2010). As opposed to regression trees in which the output of the function is necessarily a continuous numerical value, classification DTs have outputs that are discreet class labels (Joshi, 2020), which realizes one of the main advantages of most applications of DTs: modeling non-numerical data, as with the book classification graphically depicted in **Figure 21**, and the example of **Figure 22**. Nonetheless classification trees can also be used to accurately model ordinal numeric data.

Training a classification tree is essentially the same as that with regression trees; however, as we do not have numerical outputs, the pruning process cannot use the squared error. In its place, there are several measures. The misclassification measure, Gini index, and cross-entropy are the most used (Joshi, 2020), which are also used as criteriums to make decisions to split in nodes. As we have already seen, different metrics or measures to make the separation decision at each node differentiate the various decision tree algorithms which, apart from their intended use (regression or classification), in most of the literature ends up identifying the different varieties of decision trees such as ID3 trees, or CART trees, and others.

Joshi characterizes these measures as:

> *Let there be "k" classes and "n" nodes. Let the frequency of class (m) predictions at each node (i) be denoted as f_{mi}. The fraction of the classes predicted as m at node i be denoted as p_{mi}. Let the majority class at node m be c_m. Hence the fraction of classes c_m at node m would be p_{mc_m}* (Joshi, 2020)

The **misclassification measure** is the most intuitive:

> *Based on the variables defined above, the misclassification rate is defined as* $1 - p_{mc_m}$. *As can be seen from [***Figure 25***], this rate is not a continuous function and hence cannot be differentiated. However, this is one of the most intuitive formulations and is fairly popular.* (Joshi, 2020)

On the other hand, the **Gini index** is the measure used in the **CART** algorithm to build a decision three:

> *The Gini index is the measure of choice in CART. The concept of the Gini index could be summarized as the probability of misclassification of a randomly selected input sample if it was labeled based on the distribution of the classes in the given node. Mathematically it is defined as [shown in equation* (8)*].*

$$\mathcal{G} = \sum_{m=1}^{k} p_{mi}(1 - p_{mi}) \tag{8}$$

> *This is a smooth function of the proportion and is continuously differentiable, and can be safely used in optimization, as shown in [***Figure 25***].* (Joshi, 2020)

Given the advantages of smooth function and a measure of impurity, such as the Gini index that works as a probability, the CART algorithm is used and tested in later sections of this dissertation.

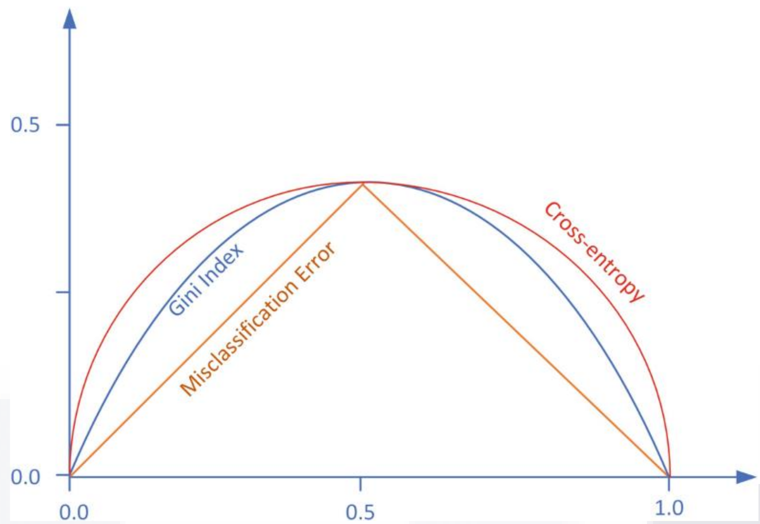


Figure 25: *The plot of decision metrics for a case of 2 class problems in which the X-axis shows the proportion in class 1; curves are scaled to fit without loss of generality* (Joshi, 2020)

Finally, **cross-entropy** is the measure used in ID3:

> *Cross-entropy is an information-theoretic metric that resembles the classical entropy of a single random variable. However, as the random variable here is already a combination of the class prediction and nodes of the tree, it is called cross-entropy. ID3 models use cross-entropy as the measure of choice. As the plot in the figure shows, this is a smooth function of the proportion and is continuously differentiable and can be safely used in optimization.* (Joshi, 2020)

Joshi (2020) presents the cross-entropy as an information-theoretic metric defined as shown in equation (9):

$$\mathcal{E} = -\sum_{m=1}^{k} p_{mi}\log_2(p_{mi}) \qquad (9)$$

IV.4.7. Training and use of decision trees

Joshi (2020) presents a general procedure for practical training decisions trees using CART or ID3:

1. Start with the training data.
2. Choose the metric of choice (Gini index or cross-entropy).

3. Choose the root node such that it splits the data with optimal values of metrics into two branches.
4. Split the data into two parts by applying the decision rule of the root node.
5. Repeat the steps 3 and 4 for each branch.
6. Continue the splitting process till leaf nodes are reached in all the branches with predefined stop rule. (Joshi, 2020)

The process described in the above steps is what we have seen as recursive portioning in its TDIDT approach, as it begins building from the top of the tree, creating nodes all the way down to the leaves. Fürnkranz (2017) characterized this process thus:

> *Decision trees are learned in a top-down fashion, with an algorithm known as top-down induction of decision trees (TDIDT), recursive partitioning, or divide-and-conquer learning. The algorithm selects the best attribute for the root of the tree, splits the set of examples into disjoint sets, and adds corresponding nodes and branches to the tree. The simplest splitting criterion is for discrete attributes, where each test has the form* $t \leftarrow (A = v_t)$ *where* v_t *is one possible value of the chosen attribute A. The corresponding set S t contains all training examples for which the attribute A has the value t. This can be easily adapted to numerical attributes, where one typically uses binary splits of the form* $t \leftarrow (A < v_t)$*, which indicate whether the attribute's value is above or below a certain threshold value* v_t *. Alternatively, one can transform the data beforehand using a discretization algorithm.* (Fürnkranz, 2017)

This discretization algorithm can be seen in **Algorithm 1**.

```
function TDIDT(S)
Input: S, a set of labeled examples.
        Tree = new empty node
        if all examples have the same class c
        or no further splitting is possible
        then // new leaf
                Label (Tree) = c
        else // new decision node
                (A, T) = FindBestSplit(S)
                for each test t ∈ T do
                        S t = all examples that satisfy t
                        Node t = TDIDT(S t )
                        AddEdge (Tree⟶tNodet)
                endfor
        endif
        return Tree
```

Algorithm 1: Discretization algorithm to create train tree; this functions outputs a whole tree instead of just a new node, but it needs to test every case (Fürnkranz, 2017)

> *After splitting the dataset according to the selected attribute, the procedure is recursively applied to each resulting dataset. If a set contains only examples from the same class, or if no further splitting is possible (e.g., because all possible splits have already been exhausted or all remaining splits will have the same outcome for all examples), the corresponding node is turned into a leaf node and labeled with the respective class. For all other sets, an interior node is added and associated with the best splitting attribute for the corresponding set, as described above. Hence, the dataset is successively partitioned into nonoverlapping, smaller datasets until each set only contains examples of the same class (a so-called pure node). Eventually, a pure node can always be found via successive partitions unless the training data contains two identical but contradictory examples: examples with the same feature values but different class values.* (Fürnkranz, 2017).

As we have seen, so long as a compatible criterion is used to create new nodes, trees using the TDIDT or recursive partitioning approach can be used for regression and classification if the output y is quantitative or qualitative, respectively.

An example of a regression tree in which y follows a normal distribution is provided by Chipman in **Figure 24**, where we can see both the tree and 800 observations of recursive partitioning.

IV.4.8. Assembles of decision trees

In practice, given their efficiency, most implementations of decision trees are ensembles, that is, sets of decision trees in which different decisions are taken at some nodes and then compared between them. Joshi (2020) characterizes them as follows:

> *Such techniques are called ensemble methods and typically deliver superior performance at the cost of computation and algorithmic complexity. In ensemble methods, a single decision tree is called a single learner or weak learner, and the ensemble methods deal with a group of such learners. There are various approaches proposed in the literature that can successfully combine multiple weak learners to create a robust overall model.*
>
> *Each weak learner in the ensemble of learners captures certain aspects of the information contained in the data that is used to train it. The job of an ensemble tree is to optimally unite the weak learners to have better overall metrics. The primary advantage of ensemble methods is a reduction in overfitting.* (Joshi, 2020)

As Joshi clarifies, weak learner and strong learner have the following meanings in this context:

> *A weak learner is a decision tree trained using only a fraction of the total data and is not capable or even expected of giving metrics close to the desired ones. The theoretical definition of a weak learner is one whose performance is only slightly better than pure random chance. A strong learner is a single decision tree that uses all the data and can produce reasonably good metrics. In ensemble methods, an individual tree is always a weak learner as it is not exposed to the full data set.* (Joshi, 2020)

As Joshi (2020) mentions, there are three main types of ensembles: bagging, random forest, and boosting. The first two are relevant to this dissertation.

IV.4.8.1. Bagging

Bagging is the simplest assemble method. It simply means building various decision trees with different random subsets of the training dataset and then averaging the results in the case of regression or taking the mode in the case of classifications. Joshi presents the following steps to perform a bagging assemble:

1. Split the total training data into a predetermined number of sets with random sampling with replacement. The term "with replacement" means that the same sample can appear in multiple sets, and each sample is called a Bootstrap sample.
2. Train decision tree using CART or ID3 method using each of the data sets.
3. Each learned tree is called a weak learner.
4. Aggregate all the weak learners by averaging the outputs of individual learners for the case of regression and aggregate all the individual weak learners by voting for the case of classification. The aggregation steps involve optimization, such that prediction error is minimized.
5. The output of the aggregate or ensemble of the weak learners is considered the final output. (Joshi, 2020)

Joshni empathizes that even as this method does not require complex mathematics, it is pretty effective and has excellent resilience to outliers because of the random sampling; only one or a few decision trees are affected.

IV.4.8.2. Random forests

One of the most used implementations of decision trees, random forest, can be defined as a "hybrid of the Bagging algorithm and the random subspace method and uses decision trees as the base classifier" (Sammut & Webb, 2017e). The random subspace method is a technique that is used to generate ensembles of many kinds of classifiers, and it is described by Sammut as follows:

> *The random subspace method is an ensemble learning technique. The principle is to increase diversity between ensemble members by restricting classifiers to work on different random subsets of the whole feature space. Each classifier learns with a subset of size n, chosen uniformly at random from the complete set of size N. Empirical studies have suggested promising results can be obtained with the rule-of-thumb to choose $n = N / 2$ features. The method is generally found to perform best when there are many features (large N) and the discriminative information is spread across them. The method can underperform in the converse situation when there are few informative features and a large number of noisy/irrelevant features. Random Forests is an algorithm combining RSM with the Bagging algorithm, which can provide significant gains over each used separately.* (Sammut & Webb, 2017e)

In the case of the random forests, each decision tree is "constructed from a bootstrap sample from the original dataset. An important point is that the trees are not subjected to pruning after construction, enabling them to be partially overfitted to their own sample of the data. To further diversify the classifiers, at each branch in the tree, the decision of which feature to split on is restricted to a random subset of size n, from the full feature set. The random subset is chosen anew for each branching point. The value for n is suggested to be $\log_2(N + 1)$, where N is the size of the whole feature set". (Sammut & Webb, 2017d)

Random forest ensembles go beyond the strengths provided by bagging ensembles to minimize the undesirable effects of interdependence of features of the dataset and their unequal influence on the outcome of a decision tree. Joshni explains how it does this:

> *[Random forest get their advantages] by [randomly] partitioning feature space as well as data for an individual weak learner. Thus each weak learner sees only a fraction of samples and a fraction of features. The features are also sampled randomly with replacement […], as data is sampled with replacement in bagging methods. The process is also called as random*

> *subspace method, as each weak learner works in a subspace of features. In practice, this sampling improves the diversity among the individual trees and overall makes the model more robust and resilient to noisy data. The original algorithm proposed by Tin Ho was then extended by Breiman […] to merge the multiple existing approaches in the literature into what is now commonly known as the random forest method.* (Joshi, 2020)

Finally, the result is chosen in the same way as it is for bagging, by averaging the outputs of the weak learners or using their mode. Given that random forests retain the simplicity of bagging methods but tend to be more accurate, these are the kinds of ensembles tested and used in this dissertation.

IV.5. Other classification algorithms used

The high performance of deep learning classification algorithms based in feedforward ANNs and the high interpretability of decision trees-based methods made them the most attractive classification algorithms for the task at hand in the research presented in this dissertation. Nonetheless, two other algorithms were included supplementally because of their characteristics that offered a benefit to their inclusion in the tests: logistic regression and support vector machines.

Firstly, logistic regression is already a highly used tool in quality-of-life studies (Das, 2014) and thus of particular interest to the themes of the research on this dissertation. Critically, it has already been tested with one of the datasets used in this dissertation and found to have an acceptable if not stellar performance, but with very low computational time usage, which makes a matter of seconds what other algorithms compared to it take several minutes or even hours (Marin & Ponce, 2020). Hence, this algorithm is also included for comparison's sake to have a reference of a method with much better computational time compared to deep learning and ensembles of decision trees.

A similar reason gained the algorithm of support vector machines its inclusion in tests. They have been found to have similar performance to deep ANNs classification algorithms for one of the datasets used in this dissertation, but with a fraction of their computational time (Marin & Ponce, 2020). While logistic regression and support vector machines serve as benchmarks in the tests in later chapters, these chapters go into these methods in a brief manner unlike with, for instance, decision trees whose characteristics and architecture are connected to the purposes and conclusions of this dissertation.

IV.5.1. Logistic Regression

Logistic regression "provides a mechanism for applying the techniques of linear regression to classification problems" (Sammut & Webb, 2017b) to fit a regression model to a dependent binary variable. As already mentioned, is a highly common tool in quality-of-life studies (Das, 2014) because it permits the study of the relation between a proportion and a set of explanatory variables, either quantitative or qualitative ("Logistic Regression," 2008), which also makes it a helpful tool in epidemiological studies as well as social sciences in general.

Logistic regression uses a regression model in the form shown in equation (10).

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (10)$$

In the above expression, $x_1$to x_n represent the values of the n attributes, and $\beta_o to\ \beta_n$ represent weights. This model is then mapped to the interval [0,1] (Sammut & Webb, 2017b). This mapping is done using the expressionshown in equation (11).

$$P(c_0|x_1, \dots x_n) = \frac{1}{1+e^{-z}} \qquad (11)$$

In the above expression, c_o represents class 0.

Logistic regression usefulness is because of two of its main characteristics:

> ***Its sigmoidal shape*** *(of the previously given expression $P(Y_j = 1)$ fits very well the relation usually observed between a dose X and the frequency of a disease Y as [expressed in equation* (12)*].*

$$\frac{e^x}{1+e^x} \qquad (12)$$

> ***It is easy to interpret:*** *the association measure between a disease and a risk factor M (corresponding to a binary variable X_i) is expressed by the odds ratio. This is a good approximation of the relative risk when the probabilities $P(Y = 1|X_i = 1)$ and $P(Y = 1|X_i = 0)$ are small, and it is computed simply as the exponential of the parameter that is*

associated with the variable X_i ; thus: the odds–ratio of the factor $M = exp(\beta_i)$ ("Logistic Regression," 2008)

The distinctive sigmoidal shape of the logistic regression can be seen in **Figure 26**.

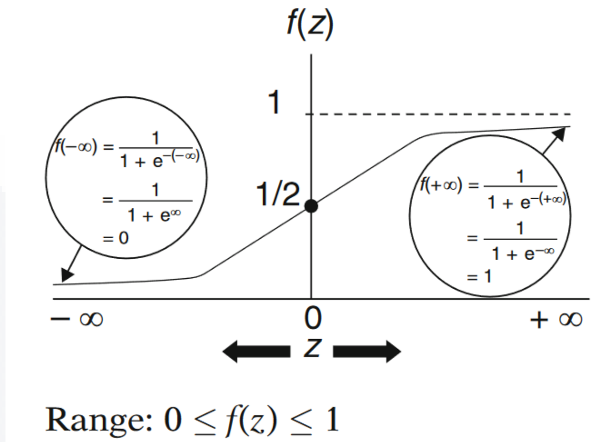


Figure 26: *Sigmoidal shape of the logistic regression (Kleinbaum & Klein, 2010)*

An instructive example is shown in the Concise Encyclopedia of Statistics in this form:

We take an example given by Hosmer and Lemeshow. The dependent variable indicates the presence or absence of heart disease for 100 subjects, and the independent variable is age. The resulting logistic regression model is [shown in equation (13)*]:*

$$\pi(X) = -5.3 + 0.11 \cdot X \qquad (13)$$

with a standard deviation for the constant and the β (coefficient of the independent variable X) equal to 1.1337 and 0.241, respectively. The two estimated parameters are statistically significant.

The risk of having heart disease increases by $e^{0.11} = 1.12$ *every year.*

It is also possible to compute the confidence interval by taking the exponential of the interval for β *[as shown in equation* (14)*]:*

$$(e^{1.11-1.96\cdot0.241}, e^{1.11+1.96\cdot0.241}) = (1.065,1.171) \qquad (14)$$

TESIS TESIS TESIS TESIS TESIS

[Table 11] gives the values for the log-likelihood as well as the parameter values that are estimated by maximizing the log-likelihood function in an iterative algorithm. ("Logistic Regression," 2008)

Table 11: *Values for log-likelihood and parameters, for example, exercise* ("Logistic Regression," 2008)

	Log-likelihood	Constant	β
1	-54.24	-4.15	0.872
2	-53.68	-5.18	1.083
3	-53.67	-5.31	1.109
4	-53.67	-5.31	1.109

Very commonly, it is the case that a model must be done to handle multiple categories instead of just two. For these cases Polytomous logistic regression is used. Most ML implementations in general programming languages include at least logistic regression with one and multiple categories using the polytomous model.

IV.5.1.1. Justification for the use of logistic regression

Logistic regression, as mentioned in the previous section, is very useful and has a low use of computational time (Marin & Ponce, 2020) for wellness and wellbeing studies as logistic regression is concerned with the particular situation in regression modeling, where the outcome is of a binary or dichotomous (yes/no) nature. Such outcomes are expected in quality of life (QOL) research (Das, 2014).

Das presents some other wellness and wellbeing measures where logistic regression is useful:

> *In addition, measures of QOL such as the Health Complaints Scale and the Global Mood Scale are frequently dichotomized (often using a median split) to create a binary outcome of "impaired QOL," for which logistic regression would be the appropriate method of statistical analysis* (Das, 2014; Denollet et al., 2000)

Kleinbaum and Klein (2010) present many reasons behind the popularity of logistic regression in practical research, among them:

- *The fact that the logistic function $f(z)$ ranges between 0 and 1 is a primary reason. In epidemiologic terms, such a probability gives the risk of an individual getting a disease. The logistic model, therefore, is set up to ensure that whatever risk estimate we get, it will always be some number between 0 and 1. Thus, for the logistic model, we can never get a risk estimate either above 1 or below 0. This is not always true for other possible models.*
- *The S-shape of $f(z)$ indicates that the effect of z on an individual's risk is minimal for low z until some threshold is reached. This threshold idea is thought by epidemiologists to apply to various disease conditions. In other words, an S-shaped model is considered to be widely applicable for considering the multivariable nature of an epidemiologic research question.* (Kleinbaum & Klein, 2010)

Some of the most important variables predicted and analyzed in this dissertation are of binary outcome, such as presence or absence of satisfaction in time use in a variety of activities.

IV.5.2. Support Vector Machines

Another classification technique that is relevant to try along with decision trees-based methods, ANNs, and logistic regression are support vector machines (SVM) methods. SVMs are mainly used for classification problems but can also be used for regression. In SVMs, as Shalev-Shwartz and Ben-David (2013) put it, data is conceptualized as points in an n-dimensional space where n is the number of data attributes. Each instance is transformed into a coordinate in said n-dimensional space. The simplest example to illustrate is the one in two dimensions (n=2) in which a set of data points are to be classified into two distinct linearly separable groups. In two dimensions, linearly separable groups can be separated by a straight line, as shown in the diagram in **Figure 27**.

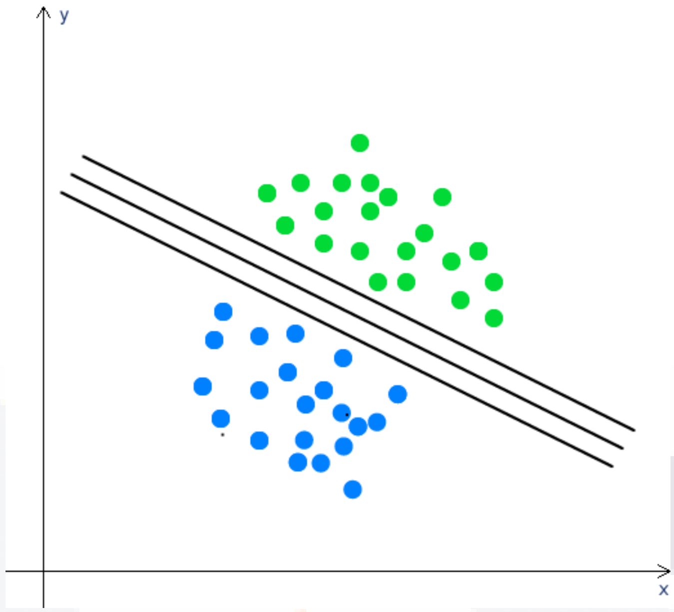


Figure 27: *Example of data classification by three different separating hyperplanes*

The three straight lines shown in **Figure 27** are known as separating hyperplanes. As Chen remarks, in an n-dimensional space, a separating hyperplane is always a (n-1)-dimensional subspace (Chen, 2018); and it can be mathematically represented as shown in equation (15).

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{n-1} x_{n-1} = 0 \qquad (15)$$

In the example shown in **Figure 27**, assuming that the label "y" is 1 (for green) or -1 (for red), all three lines are separating hyperplanes for the points of different classes. They are all separating hyperplanes because they share the same property, the data points above the line are green, and the data points below the line are red (Ray, 2015). This classification can be mathematically represented as shown in equations (16) and (17).

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 > 0 \quad \text{si} \quad y = 1 \quad \text{(green)} \qquad (16)$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 < 0 \quad \text{si} \quad y = -1 \quad \text{(red)} \qquad (17)$$

A mathematical generalization can be written as shown in equation (18).

$$y(\beta_0 + \beta_1 x_1 + \beta_2 x_2) > 0 \qquad (18)$$

The previous restriction is fulfilled only in the perfect scenario when the classes are linearly separable. An SVM model can satisfy this constraint, but in non-linearly separable scenarios, this constraint will have to be relaxed (Chen, 2018), or a data transformation will have to be used.

Given a hyperplane and a set of points, if the distance from each point to the hyperplane is calculated, all the distances will be different. The smallest distance is the margin. Figure 28 shows the distance between both sides of the dashed line to the solid line is the margin. This optimal line is the midline of the widest stretch between the red and green points (Ray, 2015).

In the linearly separable case, the SVM method tries to find the hyperplane that maximizes the margin, provided that both classes are classified correctly.

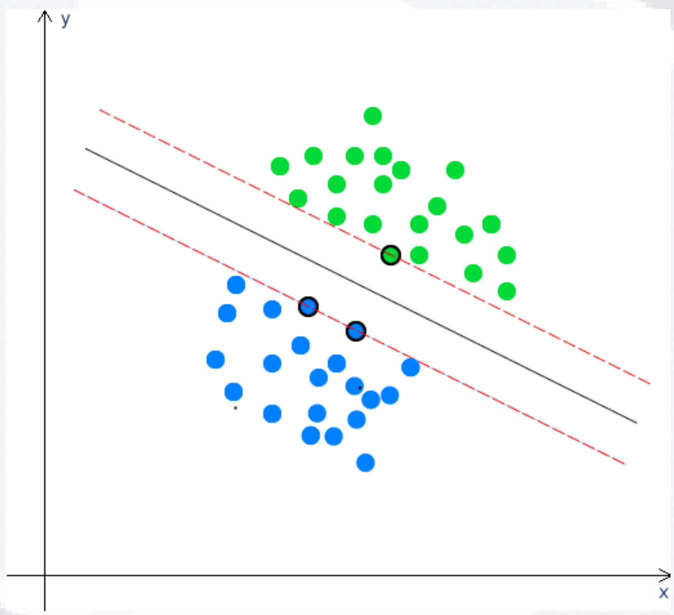


Figure 28: *Example of data classification by separating hyperplane with the margins in dashed lines and the support vectors circled*

In **Figure 28**, a handful of points are encircled; these are the support vectors that give their name to SVMs. These points are the data points closest to the separating hyperplane, the points in a dataset that, if removed, would alter the position of the separating hyperplane. Because of this, they can be considered the critical elements of a data set (Ray, 2015). We can see in **Figure 28** that if the encircled points were to disappear, the position of the separating hyperplane would be different, as the margin would maximize supported by a different set of points which, as they are contained in a 2-dimensional space, can also be thought as of 2-D vectors, hence the name support vectors.

In real-world applications, linearly separable datasets like the neat dataset shown in **Figure 28** are a veritable rarity. As we have already seen, the condition of being 100% correctly classified by a hyperplane will never be met in these cases. SVM techniques address non-linearly separable cases by introducing two concepts: soft margin and kerneling (Ray, 2015).

Figure 29 shows three data sets, of which only the first one is linearly separable, while the other two show different cases of linearly non-separable data sets.

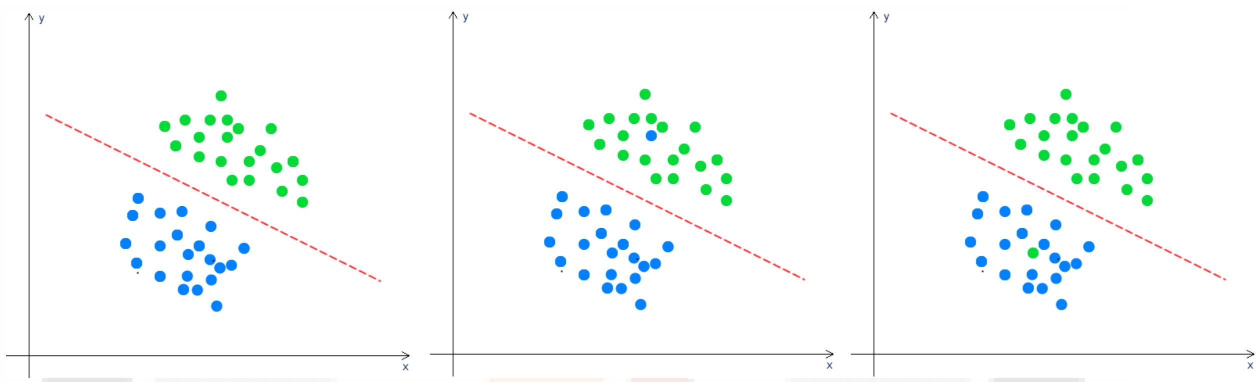


Figure 29: Three data sets; the first is linearly separable, and the second and third are not

SVM tolerates two types of misclassifications with a soft margin, which can be seen in **Figure 30**, and they are:

- A point on the incorrect side of the decision boundary but on the correct side of the margin (shown on the left of **Figure 30**).
- The point is on the incorrect side of the decision boundary and the incorrect side of the margin (shown on the right of **Figure 30**) (Ray, 2015)

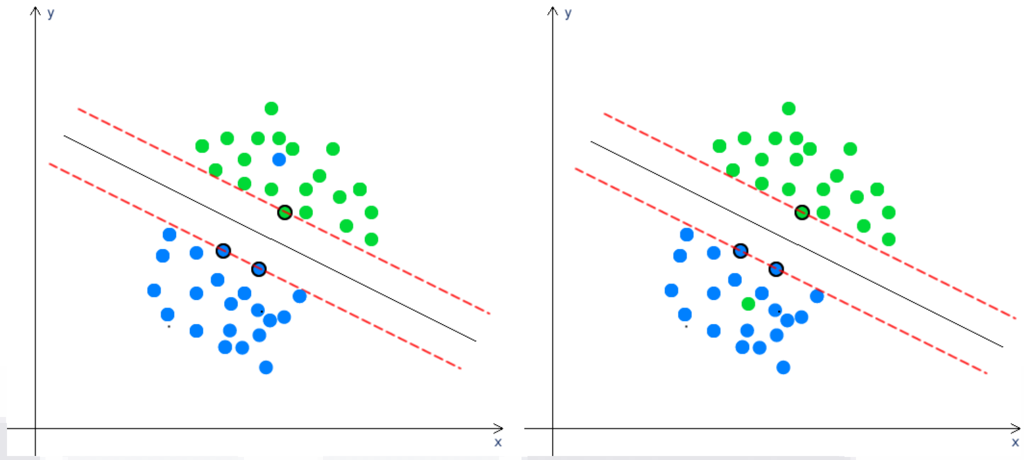


Figure 30: *Misclassifications with soft margin*

By implementing a soft margin in an SVM, the model will tolerate a relatively small number of misclassified points as it searches for a separating hyperplane that maximizes margin and minimizes misclassification. In most SVM models, this tolerance is a crucial hyperparameter. For example, in the Sklearn implementation used in this dissertation, it is represented as the penalty term “C.” The larger the C, the more penalty SVM gets when it misclassifies. Therefore, the narrower the margin and the decision limit (separating hyperplane) will depend on fewer support vectors (Ray, 2015).

Another option for taking up non-linearly separable problems with an SVM is kernelling which involves applying data transformations that create new attributes for the data points. As the data points are represented as points in an n-dimensional space, these new attributes become new dimensions in this space. These new dimensions are the key for SVM to find the nonlinear decision limit. (Chen, 2018).

Chen’s diagram in **Figure 31** shows the transformations of a data set using, first without transformation, then with a polynomial transformation, and finally with an RBF. This transformation was obtained with the help of Sklearn.

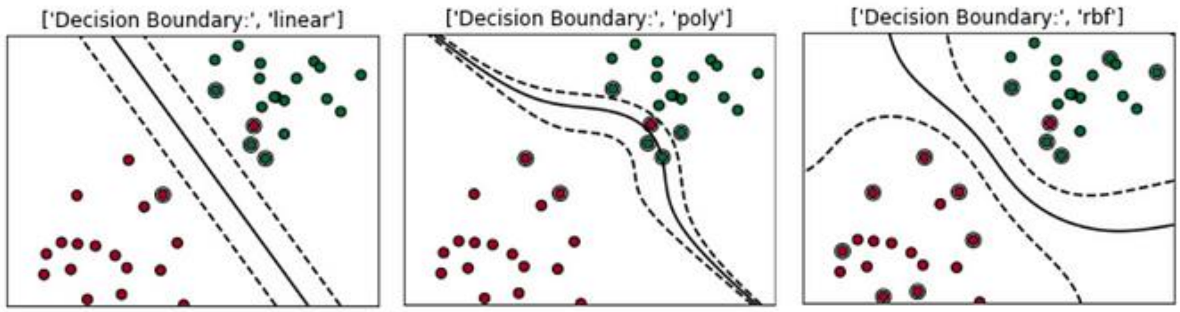


Figure 31: Dataset in its original space with separating hyperplanes, first without a transformation and then using a polynomial transformation and RBF (Chen, 2018)

It should be noted that although the separating hyperplane looks curved in the original space of the data set, it is straight in the space above where the additional dimension or dimensions created by the new attributes are included. These perspectives can seem confusing at first; Bambrick provides a small that serves to clarify them.

In the example shown in **Figure 32**, a linear separating hyperplane between the two classes cannot be had. SVM can solve this problem by adding a new attribute $z = x^2 + y^2$, creating a paraboloid in the new dimension z, with the advantage that all values in the new dimension are positive. In **Figure 33**, we can see the same data set but now plotting the x and z axes, the latter corresponding to the new attribute created as $z = x^2 + y^2$. In **Figure 33**, we must consider that all values for z would always be positive because z is the sum of the squares of x and y; and that, in the original plot, the red circles appear near the origin of the x and y axes, leading to a lower z value and a star relatively far from the origin resulting in a higher z value (Bambrick, 2016).

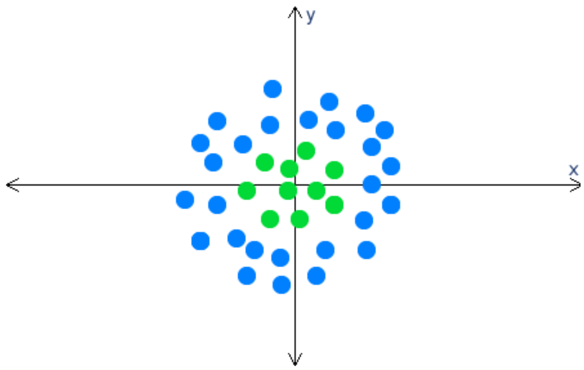


Figure 32: Example of a data set that is not linearly separable

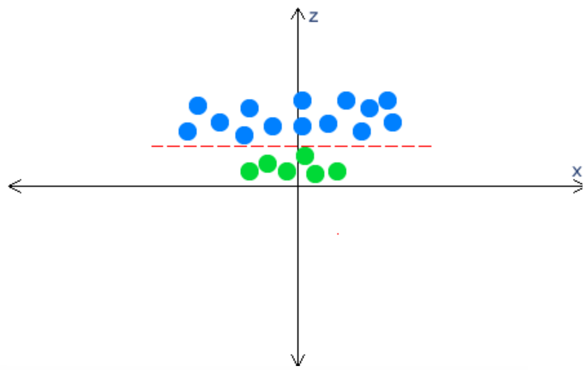


Figure 33: The same data set as ***Figure 32****, but plotting the x and z axes. A hyperplane, in this case, a 2-dimensional plane, can easily separate the points*

Thus as Bambrick (2016) remarks, when we look at the hyperplane in the original input space, it looks like a circle, as shown in **Figure 34**. However, the separating hyperplane is an entirely straight plane in the augmented space.

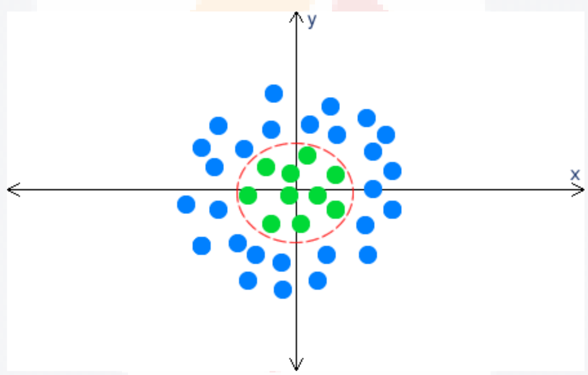


Figure 34: Dataset with separating hyperplane created in space above but viewed from the original space of the dataset

In practice, SVM implementations, such as SKLearn's, have a technique called a kernel trick. Chen (2018) describes these functions as taking a low dimensional input space and transforming it into a higher dimensional space, intending to take a non-separable problem, like the one in Figure 32, and transform it into a separable problem, like the one in **Figure 33**.

In a nutshell, procedures like the kernel trick perform some extremely complex data transformations (unlike in the example above) to figure out the process for separating data based on labels or results that have been predefined (Chen, 2018).

IV.5.2.1. Multiclass SVM

In some of the problems tackled in this dissertation, the data points must be classified into more than two classes. As we have seen in the previous subsections, SVM seems to work only for binary classification, so an iterative approach must be used to make it into a multiclass model.

According to Xanthopoulos (2013), given a total of N classes in this iterative approach, the SVM problem is solved for all binary even combinations of classes. For instance, for a three-class problem (class A, class B, and class C), we find the hyperplanes of separation that correspond to the problems A vs. B, A vs. C, and B vs. C. When a new point arrives, then each classifier "decides" on the class of this point. Finally, the point is classified in the class with the most "votes" (Xanthopoulos et al., 2013). In a sense, a multiclass SVM-based model is an SVM ensemble.

Nonetheless, Xanthopoulos (2013) warns that for the majority voting process, it is necessary to build a large number of training binary classifiers to infer the class of an unknown sample. This process can pose a computational issue in terms of performance. Therefore, in the directed acyclic graph, one tries to minimize the required necessary classifiers. This minimization can be achieved by considering a tree that removes a class at each level. An example with four classes is illustrated in **Figure 35**.

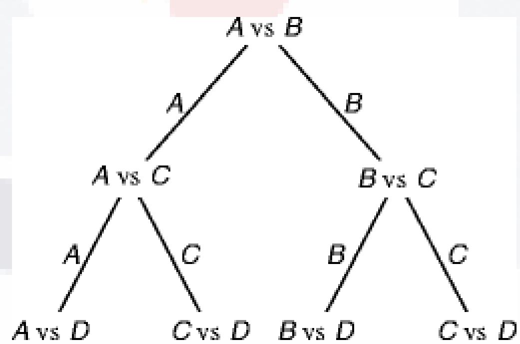


Figure 35: *Directed acyclic graph approach for an example of four classes (A, B, C, and D). At the first level of the tree, the sample goes through the A vs. B classifier. Then, depending on the result, the sample is tested through A vs. C or B vs. C. The total number of binary classifiers needed is equal to the depth of the tree* (Xanthopoulos et al., 2013)

Also, it is to be noted that multiclass problems can use the same approaches seen in the previous subsections; that is, they can use soft margins or kernelling if the datasets do not contain perfectly linearly separable classes.

IV.5.2.2. Justification for the use of SVMs

The SVM method has certain advantages and disadvantages depending on the task. However, generally, its advantages are:

- Accuracy.
- It works well on small, clean data sets.
- It may be more efficient because it uses a subset of training points. (Bambrick, 2016)

On the other hand, its disadvantages are:

- It is unsuitable for larger data sets as the training time with SVM can be high.
- Less effective on noisier data sets with overlapping classes. (Bambrick, 2016).

Since their inception, SVMs have been widely used in various applications. To give some examples, good results have been obtained in text categorization, handwritten character recognition, and bio sequence analysis. (Cristianini & Ricci, 2016). This dissertation includes them as a classification model to benchmark the other more relevant models with a very well-known and mature model.

IV.6. Cross-validation and performance metrics

Given the ML algorithms presented thus far can have different performance even while working with the same training and test datasets, there needs to be a set of performance metrics with which to compare them as well as a process to validate these measurements. Accuracy will be our main performance metric, with precision and F1 score added when relevant to check whether the accuracy metric could be misleading. Also, when relevant computational time will be measured when comparing different algorithms for practical purposes. For validation, cross-validation regimes will be used.

IV.6.1. Performance metrics

To define the performance metrics used through this text a few basic concepts need to be defined firsts in terms of predictions.

> *A true positive is an outcome where the model correctly predicts the positive class. Similarly, a true negative is an outcome where the model correctly predicts the negative class.*

A false positive is an outcome where the model incorrectly predicts the positive class. And a false negative is an outcome where the model incorrectly predicts the negative class. (Google, 2022)

Given these definitions, we can see that accuracy "is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right" (Google, 2023) and is mathematically defined as:

$$\text{Accuracy} = \frac{Correct\ predictions}{Total\ number\ of\ predictions} \tag{19}$$

Or:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

Where:

- TP: True positive
- TN: True negative
- FP: False positive
- FN: False negative

Precision, on the other hand, is the proportion of positive identifications that were correct and is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{21}$$

Finally, the F_1 score is a combination of the precision and recall metrics. Recall is a performance metric that answer the question "When ground truth was the positive class, what percentage of predictions did the model correctly identify as the positive class?" (Google, 2023) and is calculated as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{22}$$

Therefore, the F1 score can be calculated thusly:

$$F_1 = 2\frac{\text{Precision} \, * \, \text{Recall}}{\text{Precision} + \text{Recal}} \qquad (23)$$

Sometimes the accuracy metric can be misleading, so for the evaluation of the predictive power of the main training datasets, both precision and F_1 score will be added to the predictive power analysis. Even as we are actually interested in the accuracy metric provided that it is a meaningful indicator of the usefulness of the predictive classifier algorithm that is used, which is the role of the other performance metrics.

IV.6.2. Cross-validation

Cross-validation regimes are a highly regarded method to validate the performance of classifier algorithms in the ML community. These regimes are useful when the analyst or researcher has only one labeled dataset, that is, a dataset with values for the column of the variable of interest that needs to be predicted. An usual approach to evaluation the predictive power of dataset is randomly dividing the dataset in a training sub dataset and test sub dataset, usually by 80%/20% proportion respectively. But this approach has the problem than even a random selection can be non-representative.

Cross-validation is an attempt to solve this problem by dividing the dataset in k subsets, with k=10 being standard in the ML community and using each of the k sets as test datasets while the other remaining k-1 subsets remain as training dataset. More formally cross-validation can be defined thusly:

> *Cross-validation is a process for creating a distribution of pairs of training and test sets out of a single data set. In cross validation the data are partitioned into k subsets, $S_1 \ldots S_k$, each called a fold. The folds are usually of approximately the same size. The learning algorithm is then applied k times, for i = 1 to k, each time using the union of all subsets other than S_i as the training set and using S_i as the test set.* (Sammut & Webb, 2010)

In the simple case used for the test performed in this dissertation, the average of the performance metric being measured is used as the result of the cross-validation regime.

IV.7. Genetic algorithms

Genetic algorithms (GA) are a search algorithm that is classified as a metaheuristic, a class of search algorithms that generate a heuristic to search for the solution to a problem in a given search space. GA are part of evolutionary algorithms family, that are inspired in the process of evolution of living beings, and that belong to the broader family of nature inspired algorithms. GA are inspired by the process of genetic inheritance and selection and were first proposed in the seminal book "Adaptation in natural and artificial systems" (Holland, 1975) and further developed since them by different researchers and teams. Man, Tang y Kwong explain GA as follows.

> *"GA is inspired by the mechanism of natural selection where stronger individuals are likely the winners in a competing environment. Here, GA uses a direct analogy of such natural evolution. Through the genetic evolution method, an optimal solution can be found and represented by the final winner of the genetic game. GA presumes that the potential solution of any problem is an individual and can be represented by a set of parameters. These parameters are regarded as the genes of a chromosome and can be structured by a string of values in binary form. A positive value, generally known as a fitness value, is used to reflect the degree of "goodness" of the chromosome for the problem which would be highly related with its objective value."* (Man et al., 1999a)

As the data of possible solutions for a problem of changes to time allocations to a set of activities can be represented by a set of parameters, that is, the number of minutes suggested to be added or subtracted to the number of minutes already allocated to each activity considered, the problem taken up in this dissertation is amenable to GA.

> *Throughout a genetic evolution, the fitter chromosome has a tendency to yield good quality offspring which means a better solution to any problem. In a practical GA application, a population pool of chromosomes has to be installed and these can be randomly set initially. The size of this population varies from one problem to another […]. In each cycle of genetic operation, termed as an evolving process, a subsequent generation is created from the chromosomes in the current population. This can only succeed if a group of these chromosomes, generally called "parents" or a collection term "mating pool" is selected via a specific selection routine. The genes of the parents are mixed and recombined for the production of offspring in the next generation. It is expected that from this process of*

evolution (manipulation of genes), the "better" chromosome will create a larger number of offspring, and thus has a higher chance of surviving in the subsequent generation, emulating the survival-of-the-fittest mechanism. (Man et al., 1999a)

GA are by their nature iterative algorithms, that need multiple runs or "generations" to reach an acceptable solution, although as most metaheuristics they are not guaranteed to reach the optimum solution. Therefore, there need to be a termination criterion, which can be a specific level of solution's fitness, a fixed number of cycles, or other parameters of the individual solutions.

The conventional GA employed in this research has the three fundamental operators presented by Holland: crossover, and mutation. These are defined thusly:

*Crossover mechanism is depicted on [***Figure 36***]. A crossover point is randomly set. The portions of the two chromosomes beyond this cut-off point to the right are to be exchanged to form the offspring. An operation rate (P_c) with a typical value between 0.6 and 1.0 is normally used as the probability of crossover.* (Man et al., 1999a)

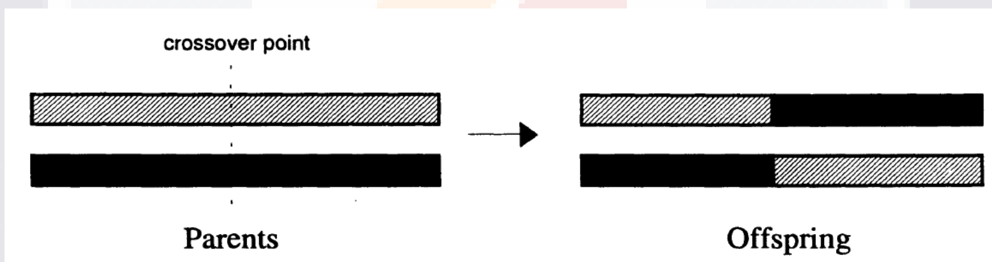


Figure 36: *Example of a one-point crossover* (Man et al., 1999a)

*However, for mutation ([***Figure 37***]), this applied to each offspring individually after the crossover exercise. It alters each bit randomly with a small probability (P_m) with a typical value of less than 0.1.* (Man et al., 1999a)

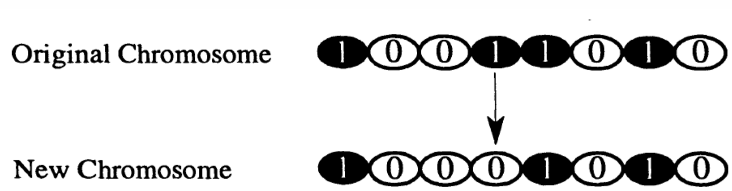


Figure 37*: Bit mutation on the fourth bit* (Man et al., 1999a)

The choice of P_m and P_c as the control parameters can be a complex nonlinear optimization problem to solve. Furthermore, their settings are critically dependent upon the nature of the objective function. (Man et al., 1999a)

While there are many modifications, operators and parameters than can be added to the conventional GA presented thus far (Man et al., 1999b), these were discovered not to be necessary for the problems tackled in the research presented in this dissertation. Henceforth any reference to GA is to the conventional algorithm presented in this section which only needs the following to be defined:

- Population
- Selection method
- Inferior limit for values
- Superior limit for values
- Probability of crossover
- Probability of mutation

V. Datasets

Given that the objective of this dissertation is to contribute evidence that the implementations of AI algorithms are an effective method to generate time management tools that contribute to time management practices as well as to wellbeing preservation and enhancement, it is of utmost importance to train such algorithms with datasets containing great quantity of datapoints of excellent quality. Furthermore, the validity of the conclusions of the experiment phase of the research presented in this dissertation depends on the quality of these training datasets, as even the nature and configuration of the data collected during the experiment depend on the information extracted from these training datasets. It is because of these reasons that the research team chose to use the microdata of the National Time Use Survey (ENUT) by the Mexican National Institute of Statistics and Geography (INEGI) aided by the National Institute for Women (INMUJERES) for training purposes.

This microdata was provided by INEGI, organized either in a big dataset file or several sub-datasets files that required a merging process. The preprocessing of the original ENUT microdata resulted in the preprocessed ENUT datasets being used as training datasets. From this point forward, any reference to preprocessed or training datasets refers specifically to clean datasets, ready to be fed to the AI algorithms, obtained from preprocessing the original ENUT microdata.

V.1. Justification of the ENUT microdata use

The ENUT is a nationwide survey with a big representative sample from all 31 Mexican states and Mexico City. Its objective is to "provide statistical information necessary for the measurement of all forms of work of individuals, both paid and unpaid, and make visible the importance of domestic production and its contribution to the economy" (INEGI, 2015). What is more, the 2019 edition of the ENUT added to the 2015 general objective that it also aims to "make visible the importance of domestic production and its contribution to the economy and, in general, the way women and men use their time, as well as the perception of their well-being…" (INEGI, 2020). These objective statements provide the central axis around which the entire design, methodology, and application of the surveys revolve and characterize the advantages and disadvantages of using the data produced by these surveys to obtain training datasets; they also shed light on why some of the best performances by the trained AI algorithms have to do with work and study.

There is a main disadvantage to using data produced by an instrument that aims to provide statistical information to measure a predefined set of metrics of interest. This disadvantage is that it is rather challenging to find new insights in this kind of data, as the concepts, constructs, and operationalizations used in endeavors as big as the ENUT are, by necessity, well studied and are known quantities to experts in the relevant fields. Time-use, work, economic and SWB constructs, and other attributes used to build the survey are so used because there is already an extensive corpus of literature and research to back their usefulness and validity to measure the country's situation and generate information that can guide public policy. Therefore, their correlates and links to other constructs or data linked to relevant research are mostly well known.

The counterpart advantage to the abovementioned disadvantage is that if any insights are to be gained from further data analysis from different perspectives or using different tools than usual, these insights are more easily judged and integrated into existing knowledge. Also, another advantage of using this data is that it is exceptionally clean, with very high statistical design standards. Finally, these datasets are naturally big, ranging from thousands to tens of thousands of complete data points with very little or no missing data.

The research team considered that given that the objective of the research presented in this dissertation resides in the analysis of the performance of tools built with AI algorithms trained with such data, the main disadvantage of using the ENUT datasets is considerably lessened in its impact, as we are not directly trying to extract new knowledge from a dataset from such a well-studied design, but are using its known properties to study the performance and effect of AI tools trained with these data. Moreover, if new insights are produced during the actions toward our objective, these are more easily explained and offered to the field experts for further study.

Given the practical advantages of data cleanliness, quantity, and excellent design in the measurement instrument that produced the data, the ENUT microdata were excellent candidates for obtaining training datasets. However, these intrinsic characteristics are only some of the reasons the research team elected to use the ENUT microdata for training purposes. The main reasons are that they are the only dataset from a highly representative sample from a nationwide high-quality survey that includes the attributes that our research needed, such as a high quantity of data related to time assigned to different activities, demographic data, and SWB data which as we have seen is linked to overall human wellbeing and, therefore, to agency.

Other options included the microdata of another nationwide survey from INEGI. One obvious candidate was the microdata produced by the National Survey of Self-Reported Wellbeing (ENBIARE), last done in 2021 in line with OECD guidelines for SWB measurement (INEGI, 2022), just as the constructs used in this dissertation. However, this survey, while including a richer set of SWB metrics and objective wellbeing ones, does contain very little time-use data. Given that the primary use for the tools to develop was to serve as intelligent aids to time management practices, this was inadequate as the great variety of activities included in the ENUT is needed for the algorithms to evaluate a user's time-use patterns better. Therefore, the ENBIARE microdata was not used.

Finally, the option of generating our own data for training purposes not only put the cart before the horse, raising the question of which data to gather before having empirical guidance about which attributes were essential to the training of the models we intended to use; but also meant overcoming obstacles of cost, time and resources, as well as the high level of difficulty to produce a representative sample of a size or quality approaching the ones used by INEGI. For example, the smallest INEGI survey whose microdata was used in this research was collected using hundreds of employees, over several months, at a cost of over 45 million Mexican pesos (INEGI, 2015) or approximately 3 million American dollars in mid-2014. Therefore, these options were shelved, and a decision was made to train with the datasets produced by preprocessing the data from the two existing samples of ENUT microdata with SWB attributes, one from the 2014 edition of the ENUT and another from its 2019 edition; and then collect a smaller amount of data specifically for an experiment for the purposes of this dissertation. Hence, in the remaining part of this section, we describe how we obtained the ENUT-2015 and ENUT-2019 training datasets.

V.2. Characteristics and measurement instrument of the ENUT-2014 microdata

The ENUT-2014 training dataset was produced by preprocessing the microdata resulting from the 2014 edition of the National Survey of Time Use (ENUT), which was applied nationwide to a representative sample in the 31 states of Mexico and Mexico City. It included not only time-use data but also some sociodemographic data and, for the first time in national surveys from INEGI, SWB data. This later addition makes the microdata from this survey an ideal candidate to produce training datasets for the research presented in this dissertation, given that each data point has more than a hundred time-use attributes linked to the SWB evaluations of the respondent.

V.2.1. Measurement instrument for the ENUT-2014 dataset

The 2014 ENUT survey was applied to completion to 42,117 Spanish-speaking respondents, not counting people whose data was collected only partially, such as young children and people unable to answer because of disability. While there is a version of this survey applied only in the context of homes in which the primary language is one of the many indigenous ones spoken in Mexico, this data requires specialized treatment for its accurate use along with the data from the Spanish-speaking population and was not employed in the creation of the preprocessed ENUT-2014 preprocessed dataset. The design of the 2014 ENUT survey for Spanish-speaking respondents can be summarized as follows:

- **Objective population:** Spanish-speaking people, at least 12 years old.
- **Effective sample:** 16,817 houses. A small minority of them harbored more than one home, family, or group of people depending on separated income streams.
- **Design type:** Probabilistic, biphasic, stratified, and by conglomerates.
- **Application type:** Face-to-face interview
- **Survey type:** Stylized.
- **Application period:** October 13 to November 28, 2014
- **References period:** People were asked about the week prior to the arrival of the survey team.
- **Measurement units:** Hours and minutes for time use. Other questions used 1-5 integer scales, Yes/No dichotomies, or Mexican pesos (MXN) for questions about money and income.
- **Geographic coverture:** All of Mexico, including both urban and rural settlements.
- **Idoneous informants:** Ideally, the head of the house or their partner would be the person giving data pertaining to the whole house or home. If that is impossible, someone 15 or older could give this data. Then all people 12 years or older would answer the whole survey, except those with a disability that could prevent this process.

The survey type is classified as stylized because it asks the respondents to estimate themselves the quantities of time that they assign to each of their activities (Committee on National Statistics National Research Council, 2000).

The survey included various sections with questions pertaining to different subjects organized into eight sections:

I. **Characteristics of the house:** Includes questions about the physical characteristics of the house, such as the kind of flooring, number of bedrooms, services present, and others.
II. **ID and characteristics of the home/family/group:** Includes questions about the person or people living in the house and the goods and services to which they have access.
III. **Sociodemographic characteristics:** Includes questions about the people living in the house, such as age, gender, relationships, access or rights to public services or other vital services, school attendance, special needs, and others.
IV. **Personal characteristics:** Includes questions about persona attributes of the people living in the house, such as language spoken, literacy, education level, ethnic self-description, and others.
V. **Work:** Includes questions about whether the people living in the house have a job, and if they do not, why, as well as questions pertaining to the characteristics of the job or the search for one.
VI. **Usual activities:** Includes questions about how much time is assigned to different kinds of activities, such as personal care, study, work, cleaning, repairs, shopping, paperwork, domestic shores, caring for other people, contributing to the community, sports, games, entertainment, socializing, watching TV or using the Internet, rest, prayer, meditation, and many others.
VII. **Subjective wellbeing:** Includes questions about the self-perception of the level of personal wellbeing pertaining to different aspects of life. This section includes questions about satisfaction and happiness.
VIII. **Unpaid activities from people not from the home/family:** Includes questions that identify homes or families that receive support and help in an unpaid manner from people, not from the same home or family.

Except for the eighth, data from each section were included in the preprocessed ENUT-2014 dataset.

INEGI employed the National Housing Framework 2012 obtained from the 2010 national census and used it as a master sample from which the houses samples for the ENUT were obtained. These samples were selected with a design that is probabilistic, stratified, monophase, and by conglomerates.

V.2.2. Preprocessing particular to the ENUT-2014 microdata

Unlike the microdata from ENUT 2019, the microdata from the ENUT 2014 was provided by INEGI in eight sub-dataset files, some with a different number of entries pertaining to one or more of the sections of the survey presented in the previous section, as can be seen in **Table 12**. For this reason, while the selection and transformation of both the ENUT 2014 and ENUT 2019 microdata are practically the same, the 2014 one involves a few extra procedures.

Table 12: *Sub-dataset descriptions of the ENUT 2014 microdata (Marin & Ponce, 2020)*

Dataset	Survey section	Data from	Entries	Features
TVivienda	I	Houses	15,058	24
THogar	II, III, VIII	Home or family units	15,500	51
TSDem	III	Sociodemographic data	56,273	7
TModulo1	IV	Time use	42,117	182
TModulo2	V, VII	Time use	42,117	192
TModulo3	VI, VII	Time use and subjective wellbeing	42,117	178
TNoResidente	VIII	Nonresidents	1,914	33
Metadatos	I-VIII	Entries	58,187	4

As we can see in **Table 12**, the TVivienda sub-dataset includes 56,273 data points; this is because it includes sociodemographic data from all the people living at a home, including those under 12 years of age and those unable to answer the survey. On the other hand, the TModulo1, TModulo2, and TModulo3 sub-datasets include only the entries related to respondents that answered the survey to completion. Other sub-datasets have fewer data points because many of these data points are common to various entries of the other bigger sub-datasets.

The preprocessing of the sub-datasets resulting from the 2014 ENUT included a series of merging and selection processes in which all the sub-datasets were sequentially merged, and a small number of attributes dropped. The preprocessing and the wellbeing analytics framework used for it were detailed in previous research (Marin & Ponce, 2020). A diagram of the merging process of the sub-dataset is shown in **Figure 38**.

Also, a single question in the 2014 survey is absent in the 2019 one. Further analysis (Marin et al., 2021) found that the attribute produced by the answer to this question was outside the set of the most relevant ones for this research and was thus eliminated during the selection process for the ENUT-2014 preprocessed dataset.

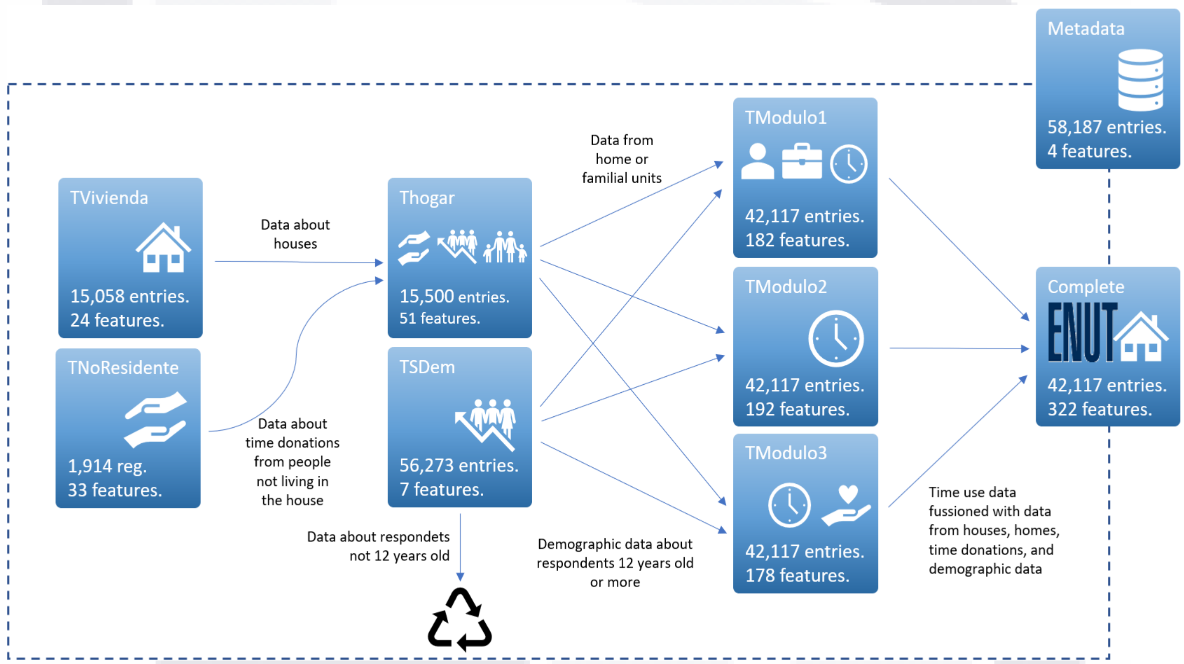


***Figure 38**: Diagram of the merging processes of the ENUT 2014 sub-datasets (Marin & Ponce, 2020)*

V.3. Characteristics and origin of the ENUT-2019 Dataset

In an analog way to the ENUT-2014 training dataset, the ENUT-2019 training dataset was produced by the preprocessing of the microdata resulting from the 2019 edition of the National Survey of Time Use (ENUT), which was applied nationwide to a representative sample in the 31 states of Mexico as well as in Mexico City. The ENUT 2019 was compatible in design with the ENUT 2014 and had only a few changes with adding a question, eliminating another, and expanding the possible answers for a couple more that were not part of the set of attributes discovered to be the most relevant for this research (Marin et al., 2021). Also, the section about unpaid activities from people not from the home/family was dropped from the ENUT 2019. However, as seen in later sections about preprocessing, this section provided no attributes for both preprocessed training datasets. Finally, for this edition of the ENUT, INEGI provided the complete microdata in a single file facilitating the preprocessing.

V.3.1. Measurement instrument for the ENUT-2019 dataset

The 2019 ENUT survey was applied to completion to 71,404 Spanish-speaking people, not counting people whose data was collected only partially, such as young children and people unable to answer because of disability. As with the ENUT 2014, a version of this survey is applied only in the context of homes in which the primary language is one of the many indigenous ones spoken in Mexico. However, this data required specialized treatment for its accurate use along with the data from the Spanish-speaking population and was not employed in creating the preprocessed ENUT-2019 dataset. Overall, the design of the 2019 ENUT survey for Spanish-speaking respondents can be summarized as follows:

- **Objective population:** Spanish-speaking people, at least 12 years old.
- **Effective sample:** 26,631 houses. A small minority of them harbored more than one home, family, or group of people depending on separated income streams.
- **Design type:** Probabilistic, stratified, and by conglomerates.
- **Application type:** Face-to-face interview
- **Survey type:** Stylized.
- **Application period:** October 21 to December 1, 2019
- **References period:** People were asked about the week prior to the arrival of the survey team.
- **Measurement units:** Hours and minutes for time use. Other questions used 1-5 integer scales, Yes/No dichotomies, or Mexican pesos (MXN) for questions about money and income.
- **Geographic coverture:** All of Mexico, including both urban and rural settlements.
- **Idoneous informants:** Ideally, the head of the house or their partner would be the person giving data pertaining to the whole house or home. If that is impossible, someone 15 or

older could give this data. Then all people 12 years or older would answer the whole survey, except those with a disability that could prevent this process.

It should be noted that all survey activities were finalized prior to any restrictions placed because of the, then incoming, COVID-19 pandemic, so no consideration of this situation is necessary.

The survey included various sections with questions pertaining to different subjects organized into seven sections:

I. **Characteristics of the house:** Includes questions about the physical characteristics of the house, such as the kind of flooring, number of bedrooms, services present, and others.
II. **ID and characteristics of the home/family/group:** Includes questions about the person or people living in the house and the goods and services to which they have access.
III. **Sociodemographic characteristics:** Includes questions about the people living in the house, such as age, gender, relationships, access or rights to public services or other vital services, school attendance, special needs, and others.
IV. **Personal characteristics:** Includes questions about persona attributes of the people living in the house, such as language spoken, literacy, education level, ethnic self-description, and others.
V. **Work:** Includes questions about whether the people living in the house have a job, and if they do not, why, as well as questions pertaining to the characteristics of the job or the search for one.
VI. **Usual activities:** Includes questions about how much time is assigned to different kinds of activities, such as personal care, study, work, cleaning, repairs, shopping, paperwork, domestic shores, caring for other people, contributing to the community, sports, games, entertainment, socializing, watching TV or using the Internet, rest, prayer, meditation, and many others.
VII. **Subjective wellbeing:** Includes questions about the self-perception of the level of personal wellbeing pertaining to different aspects of life. This section includes questions about satisfaction and happiness. Except for the eighth, data from each section were included in the preprocessed ENUT-2014 dataset.

It should be noted that, even as the ENUT 2019 uses a compatible measurement instrument with the ENUT 2014, the 2019 edition section VIII was eliminated altogether.

V.3.2. Preprocessing particular to the ENUT-2019 microdata

The microdata from the ENUT 2019 was provided in a single dataset file, greatly facilitating its handling and use along with the ENUT 2014 merged microdata file, both of which then underwent a parallel process of further preprocessing to create clean training datasets. As mentioned, some minor changes were present in the ENUT 2019 compared to the 2014 ENUT: one question was added, another was eliminated, and a couple of questions had their options for answers expanded. As preprocessing particular to the ENUT 2019 microdata, the question added to the 2019 survey was eliminated for consideration for the training dataset as well as the answers not found in a couple of the 2014 survey whose values were consolidated into the "Other option" or "No answer" equivalent.

V.4. Preprocessing of ENUT microdata

The preprocessing of the ENUT microdata began by merging all the sub-dataset files from the ENUT 2014 microdata into one table, as already shown. On the other hand, the ENUT 2019 microdata was already provided by INEGI in a single integrated microdata file. After this procedure, two datasets ready for further preprocessing were obtained, one from the ENUT 2014 and another from the ENUT 2019. From this point on, the same preprocessing was performed in both datasets.

V.4.1. Evaluation and review of the data and its encoding

A thorough evaluation of both the merged ENUT 2014 and the ENUT 2019 microdata files showed that while it included many empty values or cells, there were no missing values representing lost information. That is, in no case a respondent included in the microdata provided to the public by INEGI failed to acknowledge a relevant question. In the few cases in which the survey allows for a non-answer, for example, refusing to provide income information, that is what the empty value represents, even if, in most of these cases, a numeric code is used for that situation. However, there were thousands of empty values representing true zeros or the irrelevance of the question to the respondent. All these empty values needed to be eliminated to have functioning training datasets.

As mentioned, other minor problems were two questions with slightly different but largely overlapping sets of possible answers in the datasets from 2014 and 2019. In both cases, the very few data points with the non-overlapping values had this value integrated into the "Other" code or transformed into a zero representing this "other" option. Other than that, the quality of the microdata was outstanding.

The measurement instrument used different set of values to encode answers for each of the different kinds of questions as follows:

Binary answers

$$x_B \in \{1,2\} \tag{24}$$

In this subset, 1 encodes "Yes" and 2 "No".

Binary, with two affirmatives

$$x_{BO} \in \{1,2,3\} \tag{25}$$

In this subset, both 1 and 2 encode "Yes", and 3 for "No". The difference between 1 and 2 is:

- 1: There are one or more people with the characteristics mentioned in the question who are not the respondent, whether the respondent himself shares this characteristic or not.
- 2: There is only one person with the characteristic mentioned in the question, and it is the informant.
- 3: There is no person with the characteristic mentioned in the question.

Binary, with an option for ignorance

$$x_{BI} \in \{1,2,9\} \tag{26}$$

This subset works the same as the binary, with 1 encoding "Yes" and 2 encoding "No" but adds the option of a 9 encoding "Does not know."

Gender

$$x_G \in \{1,2\} \tag{27}$$

In this subset, 1 encodes "Male," and 2 encodes "Female".

Non-ordinal options

$$x_O \in \{1,2,4,5,6,7,8,9\} \tag{28}$$

In this subset, each number represents an option. These options are not ordered, nor do they have a greater or lesser value depending on the number that represents them. These values can represent resources available in housing or home, kinship, rights, civil status, schooling, and others.

Working hours

$$x_{HT} \in \{0,1,\dots,96,97,98,99\} \tag{29}$$

This subset represents a number of working hours per week between 0 (less than 1 hour) and 96 hours. However, the last three values in the subset have specific non-numeric values:

- 97: I did not work last week. Note that 0 hours of work is not equivalent to not working since there is also a variable for minutes of work.
- 98: 98 or more hours of work.
- 99: Does not know.

Activity hours

$$x_{HA} \in \{0,1,\dots,96\} \tag{30}$$

This subset represents a number of weekly activity hours between 0 (less than 1 hour) and 96 hours.

Minutes of work or activity

$$x_M \in \{0,1,\dots,59,97,99\} \tag{31}$$

This subset represents a number of active minutes per week that can be anywhere from 0 (less than 1 hour) to 59 minutes. These amounts can be in addition to a value of hours that is added to this amount of time or if the time dedicated to the activity or work is less than one hour.

- 97: I did not work last week. Note that 0 hours of work is not equivalent to not working since there is also a variable for minutes of work.

- 99: Does not know.

Age

$$x_E \in \{0, 1, \dots, 96; 97, 98, 99\} \tag{32}$$

This subset represents a number of years of age that can be between 0 years and 96 years. However, the last three values in the subset have specific non-numeric values:

- 97: 97 years or older.
- 98: Age is not known exactly in people 12 years and older.
- 99: Age is not known exactly in people under 12 years of age.

Money for income

$$x_D \in \{0, 1, \dots, 97999; 98000, 99999\} \tag{33}$$

This subset represents an amount of money in Mexican pesos between 0 and 97999 pesos. The remaining values have meanings as follows:

- 98000: 98000 Mexican pesos or more.
- 99999: No response.

Money, from or for contributions

$$x_D \in \{0, 1, \dots, 999998, 999999\} \tag{34}$$

This subset represents an amount of money in Mexican pesos that can be between 0 pesos and 999,998 pesos. The remaining value has a meaning as follows:

- 999999: Does not know.

Periods

$$x_P \in \{1, 2, 3, 4\} \tag{35}$$

This subset of data codes different periods in which income is received. The numbers represent:

- 1: Weekly.
- 2: Biweekly.
- 3: Monthly.
- 4: Annually.

Small amount

$$x_C \in \{1, \dots, 99\} \tag{36}$$

This subset of data encodes small amounts of resources or people as an integer number between 1 and 99.

Satisfaction (ordinal)

$$x_S \in \{1,2,3,4,5\} \tag{37}$$

This subset of data encodes a state of perceived satisfaction of the respondent:

- 1: Not at all satisfied
- 2: A little satisfied
- 3: More or less satisfied
- 4: Satisfied
- 5: Very satisfied

Satisfaction with time use

$$x_{ST} \in \{1,2,3,8\} \tag{38}$$

This subset of data encodes the level of perceived satisfaction with the time spent on a particular activity:

- 1: You would like to spend less time on it.
- 2: You are satisfied with the time spent.
- 3: You would like to spend more time on it.
- 8: Does not apply, or the activity is not carried out.

Happiness

$$x_F \in \{1,2,3,4,5\} \tag{39}$$

This subset of data encodes a state of perceived happiness of the respondent:

- 1: Not happy at all
- 2: A little happy
- 3: More or less happy
- 4: Happy
- 5: Very happy

Metadata

Some attributes were linked to survey metadata; however, as the selection process will not include them, their encoding became irrelevant for the purposes of this work.

V.4.2. Selection and single value transformations

The selection criterium for the ENUT microdata was maximalist, by selecting all the attributes included in the microdata tables from 2014 and 2019, and thus only discarding the handful of attributes included in one of them but not in the other, as well as the attributes linked to metadata in both. Then a set of single-value transformations was done by converting to zeros empty values and values representing non-answers to questions posed and acknowledged by the respondent.

Transforming all empty values into zeros is justified because, after a thorough review, there are only two cases of an empty value in the ENUT microdata:

- The question was not posed to the respondent because it was not relevant. For example, questions about paid work to a respondent who has already said they have no such activities. In this case, a zero represents the non-relevance of the question, as this is valuable information.
- Time-use assignment questions for which the respondent said they did not perform the activity in question or were assigned zero minutes or hours during the last week. In this case, a zero represents the actual answer of zero minutes.

On no occasion did any of the two cases presented above coincide in a single attribute. Thus, converting all empty values into zeros was the most straightforward mean of eliminating the problem of empty values, which cannot be fed to the algorithms, while preserving the information encoded by the presence of these empty values.

Also, all values representing a non-answer to a question posed and acknowledged by the respondent were transformed into zeros to avoid skewing the distribution of the values of the attributes including them, given that, as seen in the previous section, most of these values are on the higher end of the range for these attributes.

V.4.3. Ad-hoc transformations

To facilitate data processing, data transformations are proposed that include a change of the data from the original subspace to one more convenient for its analysis by means of transformation functions.

The design criterium of these functions is the similarity between the original and new subspaces, both in terms of possible values and the number of possible values. The attributes subsets presented in the previous section are divided into two types: those that require an independent transformation that does not need to consider the variability or range of the values and those that do. **Table 13** shows the first kind of transformation—all transformation output integer values.

Table 13: *Transformations for the subsets of attributes that do not require to consider the variability or range of the values*

Subset	Original subspace	Transformation function	Target subspace
B "Binary"	$x_B \in \{1,2\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_B = 2 \\ 1 & \text{si } x_B = 1 \end{cases}$	$x_b \in \{0,1\}$
B$_O$ "Binary, with 2 affirmatives"	$x_{BO} \in \{1,2,3\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_B = 2,3 \\ 1 & \text{si } x_B = 1 \end{cases}$	$x_{bo} \in \{0,1\}$
B$_I$ "Binary with an option for ignorance."	$x_{BI} \in \{1,2,9\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_B = 2 \\ 1 & \text{si } x_B = 1 \end{cases}$	$x_{bi} \in \{0,1\}$
O "Options"	$x_O \in \{1,2,4,5,6,7,8,9\}$	$I(x_O) = x_O$	$x_o \in \{1,2,4,5,6,7,8,9\}$
P "Periods"	$x_P \in \{1,2,3,4\}$	$I(x_P) = x_P$	$x_p \in \left\{4,2,1,\frac{1}{12}\right\}$
S "Satisfaction"	$x_S \in \{1,2,3,4,5\}$	$I(x_S) = x_S$	$x_s \in \{1,2,3,4,5\}$

ST "Satisfaction with time use."	$x_{ST} \in \{1,2,3,8\}$	$f(x_B) = \begin{cases} -1 & \text{si } x_{ST} = 1 \\ 0 & \text{si } x_{ST} = 2 \\ 1 & \text{si } x_{ST} = 3 \end{cases}$	$x_{st} \in \{-1,0,1\}$
F "Happiness"	$x_F \in \{1,2,3,4,5\}$	$I(x_F) = x_F$	$x_f \in \{1,2,3,4,5\}$
M "Minutes of activity."	$x_M \in \{0,1,\dots,59\}$	$f(x_M) = \begin{cases} 0 & \text{si } x_M = 0 \\ 1 & \text{si } 0 < x_M \leq 12 \\ 2 & \text{si } 12 < x_M \leq 24 \\ 3 & \text{si } 24 < x_M \leq 36 \\ 4 & \text{si } 36 < x_M \leq 48 \\ 5 & \text{si } 48 < x_M < 60 \end{cases}$	$x_m \in \{0,1,2,3,4,5\}$

The justifications for the transformation shown in **Table 13** are mostly practical, that is, to be able to be fed to the algorithms as training datasets, and are as follows:

B "Binary": The negative is coded with 0 and the positive with 1.

B_O "Binary with two affirmatives": This encoding is exclusive to questions about the presence of people with conditions that require special care in the household. The positive is coded only in the case of the option, which means that there is another person who is not the respondent with a condition that requires special care and to whom the respondent may have spent time treating. Since the other affirmative is that the respondent himself is the one with the condition with special care requirements, but the questions where this subset of answers is used deal with care dedicated to third parties or not to himself, the positive can be recorded, which includes only the respondent as a negative, along with the other negative both code: "No, there is no person other than the respondent who has required special attention from the respondent."

B_I "Binary with an option to ignore": The negative is coded with 0 and the positive with 1. The value corresponding to "Does not know" is also encoded as a 0.

O "Options," S "Satisfaction," F "Happiness": The identity function is used as a transformation function since these spaces can be used as is.

P "Periods": This variable is not used directly but is multiplied by the income variable to obtain a monthly income. Therefore, if it was indicated that the income was weekly, it is multiplied by 4, biweekly by 2, monthly by 1, and annual by 1/12.

S_T "Satisfaction with time spent": Desiring to spend less time on an activity with a negative value (-1) and wanting to spend more time on an activity with a positive value (1) is coded, while satisfaction with the current time spent with a neutral (0).

M "Minutes of activity": Since the minutes of activity only vary between 0 and 59, a value between 1 and 5 is assigned to fixed intervals within the possible values.

The remaining attributes not included in **Table 13** are attributes that need rescaling, and therefore there is a need to consider the range of the attribute using the following parameters:

$\{a_1, a_2, a_3, a_4\}$: Parameters for the intervals in attributes of time.

$\{b_1, b_2, b_3, b_4\}$: Parameters for small quantities attributes.

$\{c_1, c_2, c_3, c_4\}$: Parameters for money contributions attributes.

$\{d_1, d_2, d_3, d_4\}$: Parameters for money (income) attributes.

$\{e_1, e_2, e_3, e_4\}$: Parameters for age attributes.

Given the discrete nature of the datasets, the most convenient way to assign values to these parameters for our purposes was using the quintiles of the distribution for the values of each attribute. Other possibilities were dividing the range into five equally big subranges or assigning a value justified in expert knowledge to the parameters. The transformation functions that use these parameters can be seen in **Table 14**.

Table 14: *Transformations for the subsets of attributes that do require to consider the variability or range of the values*

Subset	Original subspace	Transformation function	Target subspace
H "Hours of activity."	$x_{HA} \in \{0,1,\dots,99\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_{HA} = 0 \\ 1 & \text{si } 0 < x_{HA} \leq a_1 \\ 2 & \text{si } a_1 < x_{HA} \leq a_2 \\ 3 & \text{si } a_2 < x_{HA} \leq a_3 \\ 4 & \text{si } a_3 < x_{HA} \leq a_4 \\ 5 & \text{si } x_{HA} > a_4 \end{cases}$	$x_{ha} \in \{0,1,2,3,4,5\}$
E "Age"	$x_E \in \{0,1,\dots,96;97,98,99\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_E = 0 \\ 1 & \text{si } 0 < x_E \leq e_1 \\ 2 & \text{si } e_1 < x_E \leq e_2 \\ 3 & \text{si } e_2 < x_E \leq e_3 \\ 4 & \text{si } e_3 < x_E \leq e_4 \\ 5 & \text{si } x_E > e_4 \end{cases}$	$x_e \in \{0,1,2,3,4,5\}$

D "Income"	$x_D \in \left\{ \begin{matrix} 0,1,\dots,97999;98000 \\ ,99999 \end{matrix} \right\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_E = 0 \\ 1 & \text{si } 0 < x_E \leq d_1 \\ 2 & \text{si } d_1 < x_E \leq d_2 \\ 3 & \text{si } d_2 < x_E \leq d_3 \\ 4 & \text{si } d_3 < x_E \leq d_4 \\ 5 & \text{si } x_E > d_4 \end{cases}$	$x_d \in \{0,1,2,3,4,5\}$
Dc "Money, from or to contributions."	$x_D \in \left\{ \begin{matrix} 0,1,\dots,999998, \\ 999999 \end{matrix} \right\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_E = 0 \\ 1 & \text{si } 0 < x_E \leq c_1 \\ 2 & \text{si } c_1 < x_E \leq c_2 \\ 3 & \text{si } c_2 < x_E \leq c_3 \\ 4 & \text{si } c_3 < x_E \leq c_4 \\ 5 & \text{si } x_E > c_4 \end{cases}$	$x_{dc} \in \{0,1,2,3,4,5\}$
C "Small quantity."	$x_C \in \{1,\dots,99\}$	$f(x_B) = \begin{cases} 0 & \text{si } x_E = 0 \\ 1 & \text{si } 0 < x_E \leq b_1 \\ 2 & \text{si } b_1 < x_E \leq b_2 \\ 3 & \text{si } b_2 < x_E \leq b_3 \\ 4 & \text{si } b_3 < x_E \leq b_4 \\ 5 & \text{si } x_E > b_4 \end{cases}$	$x_c \in \{0,1,2,3,4,5\}$

V.4.4. One-hot encoding transformations

One the most useful transformations in preprocessing dataset to train ML algorithms is one-hot encoding in which a categorical variable is transformed into as many binary dummy variables as different values the original categorical variable has. Given that the ENUT microdata contains many such categorical variables, this transformation was applied to all of them, except to those that are already binary. While this greatly expanded the number of attributes for the preprocessed ENUT datasets, it is a necessary procedure not only to make possible the training of the ML algorithms, but also to create dummy variables that can serve as binary variables of interest.

As a result of the preprocessing, two big training datasets suitable for training the AI algorithms used in this dissertation are produced: the ENUT-2014 and the ENUT-2019 preprocessed datasets with 42118 and 71404 complete integer data points, respectively, each of them with 432 attributes, no missing values, and fulfilling all the prerequisites to be used as training datasets for the AI algorithms presented in section IV.

V.5. Descriptive analysis and usefulness of the datasets

The ENUT datasets' value is in the information they provide about how the population of the country lives and evaluates their life in terms of SWB, particularly in terms of paid and unpaid work with a gender perspective, which is why the ENUT surveys used in this dissertation were designed in a cooperative effort between the INEGI and INMUJERES. In other words, the intrinsic value of the datasets is in their descriptive power.

However, regarding the value of the data as training datasets, it is fair to say that such value is in the prescriptive and predictive power of the datasets. They are useful for training AI algorithms that can accurately predict specific values of interest for a given sample not included in the training dataset. Nonetheless, before going into the development of AI tools part of this dissertation, it is valuable to perform a small descriptive analysis of the datasets, at least in terms of what previous research (Marin et al., 2021) has unveiled as the most critical attributes in terms of satisfaction with time use in academic and work environments, which is the most important measure for evaluating time management tools, as well as two general subjective evaluations such as general satisfaction and happiness.

The following tables in this section describe the average values for some key attributes for each possible value of an evaluation of absence or presence of satisfaction in time-use evaluations. The attributes for which the averages are computed are relevant to predict time use satisfaction in the two environments in which the tools developed and tested in this dissertation have their more likely application: work and study environments. Also, the same is done for each value of general satisfaction and happiness attributes. The ENUT-2014 dataset was used to produce the values.

Firstly, we look at the relationship between crucial SWB metrics and important environmental and personal data, such as the number of rooms and inhabitants in a respondent house, as seen in **Table 15**.

Table 15: *Average values of the number of rooms in the respondent's house, inhabitants in the house, and age of the respondent, for each value of key SWB attributes.*

	Time use satisfaction in academic activities			Time use satisfaction in work activities	
	No	Yes		No	Yes
Rooms	3.86	4.13		3.87	3.92
Inhabitants	4.46	4.83		4.53	4.47
Age	40.25	18.87		36.98	38.4
	General satisfaction				
	1	2	3	4	5
Rooms	3.58	3.37	3.7	3.89	4.15
Inhabitants	4.18	4.48	4.59	4.53	4.37
Age	45.65	42.61	39.64	37.24	34.81
	Happiness				
	1	2	3	4	5
Rooms	3.69	3.38	3.64	3.89	4.09
Inhabitants	4.22	4.49	4.53	4.52	4.48
Age	49.01	43.5	41.63	37.73	33.84

The data in **Table 15** shows that in terms of the number of rooms in a house, a proxy for size, the more rooms, the higher the SWB levels. The same can be said for the number of people living in the house. On the other hand, the trend in the relationship between age and SWB levels is inverse: in general, younger people have higher levels of SWB than older people. Previous research (Marin et al., 2021) indicated that these attributes were significant in predicting satisfaction with time use, even as they are attributes that play no part in suggestions an intelligent time management tool can make. On the other hand, attributes of time-use assignations can play a part in such suggestions, as can be inferred by the data shown in **Table 16**.

Table 16: *Average values of minutes assigned on weekdays (except as otherwise noted) for critical activities relevant for predicting time use satisfaction for each value of time use satisfaction in academic and work activities*

	Time use satisfaction in academic activities.		Time use satisfaction in work activities.	
	No	Yes	No	Yes
T. Work	1429.41	417.38	856.99	2098.67
T. Sleep	2267.57	2341.85	2306.05	2224.86
T. Sleep*	940.29	1010.54	961.85	926.77
T. Eat	347.93	361.79	353.78	342.37
T. Eat*	148.66	156.11	149.89	149.14
T. Personal care	247.36	288.07	257.13	244.46
T. Personal care*	99.47	115.06	102.99	98.76
T. Studying	132.48	1646.11	435.91	132.67
T. Studying*	34.13	385.61	103.04	37
T. Commute	17.83	217.76	56.88	19.72
T. Cooking	211.34	88.97	217.85	155.09
T. Serving food	122.97	72.9	130.79	90.63
T. Cleaning	183.81	121.52	198.45	134.68
T. Cleaning*	65.43	48.21	68.73	53.18
T. Folding	31.19	21.82	33.5	23.6
T. Securing home	22.58	16.33	20.92	23.32
T. Sports	64.88	149.2	77.26	73.09
T. Watch video/TV	282.14	264	288.05	264.86
T. Social net./Chat	81.81	239.5	108.47	90.8
T. Social net./Chat*	31.69	98.89	43.39	34.91
T. Reading	28.74	90.9	39.42	31.98

*Only on weekends

The all-important time use satisfaction metrics can be shown to have interesting relationships with the number of minutes assigned to the set of activities included in **Table 16**. Most importantly, the presence of time-use satisfaction in work and study-related activities seems strongly linked to dedicating a sizeable part of the day to such activities, suggesting that the people satisfied with not assigning time to work or study are a minority. Other interesting information gleaned from the data is that satisfaction in time-use for academic activities is inversely linked to the amount of time dedicated to cleaning and securing the house, and folding clothes, while positively linked to using social networks and reading. The time dedicated to sports plays an important role in the satisfaction of time dedicated to study, but not the time dedicated to work. Interestingly, the time dedicated to sleep does not significantly affect time-use satisfaction in either area. Similar patterns can be seen for both satisfaction with life in general and happiness level, as shown in **Table 17** and **Table 18**.

Table 17: *Average values of minutes assigned on weekdays (except as otherwise noted) for critical activities relevant for predicting time use satisfaction for each value of satisfaction with life in general*

	Satisfaction with life on general levels				
Activity	1	2	3	4	5
T. Work	1141.29	1220.4	1316.13	1307.2,	1271.41
T. Sleep	2206.63	2279.04	2270.57	2283.63	2264.55
T. Sleep*	899.83	927.61	940.57	952.69	953.12
T. Eat	302.27	339.48	335.95	350.31	366.15
T. Eat*	126	139.63	142.18	149.86	159.49
T. Personal care	222.33	233.84	243.44	251.17	272.07
T. Personal care*	89.66	92.76	98.64	100.86	100.86
T. Studying	199.49	177.49	212.27	337.31	449.28
T. Studying*	43.8	39.61	47.69	81.5	114.76
T. Commute	24.44	26.97	28.34	45.2	58.17
T. Cooking	192.93	197.86	211.78	190.92	194.38
T. Serving food	107.09	113.18	126.81	113.27	118.05
T. Cleaning	157.56	164.78	186.39	171.38	182.79
T. Cleaning*	59.95	59.07	66.48	62.41	63.4
T. Folding	33.18	29.29	31.95	28.89	31.7
T. Securing home	16.44	19.14	20.72	21.75	23.66
T. Sports	50.9	51.05	57.7	74.33	105.29
T. Watch video/TV	283.9	287.44	277.62	282.57	270.86

T. Social net./Chat	46.09	71.12	81.32	100.52	137.98
T. Social net./Chat*	18.33	28.88	30.97	40.37	53.33
T. Reading	20.79	23.59	27.65	35.69	53.22

*Only on weekends

Regarding satisfaction with life in general, the data in **Table 17** reveals that most activities that are important to time use satisfaction are not that important to general satisfaction levels, with a few exceptions: studying, commuting, sports, social network use, and reading.

Table 18: *Average values of minutes assigned on weekdays or weekends for critical activities relevant to predicting time use satisfaction for each value of happiness*

	Happiness levels				
Activity	1	2	3	4	5
T. Work	1004.32	1050.9	1281.06	1339.18	1258.65
T. Sleep	2244.15	2328	2283.68	2274.52	2273.98
T. Sleep*	916.02	945.31	942.04	948.87	955.48
T. Eat	273.45	331.44	332.08	348.19	365.88
T. Eat*	110.09	133.12	139.39	149.5	158.02
T. Personal care	221.73	232.35	235.71	248.88	272.24
T. Personal care*	86.74	97.17	97.17	100.12	108.65
T. Studying	150.45	176.38	210.44	311.84	445.63
T. Studying*	37.05	32.72	44.99	76.53	110.74
T. Commute to study	17.67	23.17	28.32	43.03	56.2
T. Cooking	230.06	207.32	201.17	191.6	198.31
T. Serving food	127.68	124.9	118.71	112.6	121.9
T. Cleaning	209.74	175.63	176.17	170.07	186.24
T. Cleaning*	71.52	61.45	63.14	62.58	64.52
T. Folding	37.49	33.67	29.97	28.64	32.13
T. Securing home	17.18	18.96	20.36	21.59	23.3
T. Sports	43.98	42.27	54.54	71.96	99.36
T. Watch video/TV	296.93	270.21	285.76	281.4	273.93
T. Social net./Chat	34.69	49.88	68.03	98.9	134.61

T. Social net./Chat*	16.76	20.05	26.57	39.82	51.82
T. Reading	14.03	15.61	24.53	35.75	48.32

*Only on weekends

Based on the data shown in **Table 18**, happiness levels are positively correlated to the same activities shown to be to general satisfaction levels, with the addition of the time assigned to eating meals and an inverse correlation with cleaning and folding clothes.

While the data presented in the tables of this section offers some information about activities and their relationships to SWB levels, they describe averages of a very diverse sample of people. They can be challenging to transform into actionable effective suggestions for a particular person. Also, as correlation does not indicate causation, this last relationship between time use and SWB can go either way. Some people allocate specific amounts of time to particular activities and have higher SWB levels, or people having higher SWB levels allocate specific amounts of time to these activities. Also, there might be conditional relationships between the allocations of time. For example, commuting seems to be positively correlated with higher levels of SWB, contrary to what research on commuting times and satisfaction indicates (Lorenz, 2018). This situation could be because commuting is a proxy for having somewhere to go or having something to do, particularly in this case, studying, which is correlated with higher SWB levels.

It is in the complexity of the possible relationships between the attributes contained in the ENUT datasets and their impact on SWB levels that the approach of using AI tools can facilitate the task of taking the information contained in these datasets from a descriptive to a prescriptive usefulness, which can be employed to actually suggest to a particular user not contained in the training ENUT datasets ways of possibly enhancing their SWB levels by modifying their patterns of time use.

VI. Development of intelligent time management tool

The research presented in this dissertation is constructed on the research team's conclusions regarding the approach to developing an intelligent time management tool that suggests changing activities considering the criteria of preserving or improving the SWB levels of users, as well as their performance. As already seen in the first chapters of this text, SWB includes emotional responses and satisfaction judgments in general and in different domains of life (Diener et al., 1999). As shown in Chapter II, the fulfillment of this criterion can be achieved because there is ample evidence that time management is linked to SWB and improved performance in work and academic activities (Aeon & Aguinis, 2017) even though there are relatively few applications and literature focused on this critical intersection, as seen in Chapter III.

VI.1. Justification of the use of ML algorithms

From the very beginning of the development of the intelligent time management tool, the research team was faced with a dilemma: to focus on the design of an expert system that would integrate a knowledge base with the principles and activities related to the preservation and increase of SWB and that, by collecting and processing some user data, would present potentially actionable information through a user interface (Hadi, 2011); or, focus on the study of massive data sets on time use and SWB, and use ML tools to predict which set of changes in time-use is most beneficial for a given user.

SWB variables relevant to the team's objective are not directly linked to the activities carried out in the day-to-day of people as such but to the relationship of these activities with a personal evaluation according to a life project whose description is linked to the culture and society in which the person lives (Şimşek, 2009). Previous research (Marin & Ponce, 2020), as well as predictive power analysis presented later in this dissertation, found that using regression tools and ML on the ENUT preprocessed datasets results in predictions that are little better than a random classifier for all measures of SWB except for satisfaction in the domain of the instrument we are developing: use of time in different types of activities, where approximately 80% and 95% of the predictions were correct for academic and work contexts respectively. These results mostly back the conclusion of the review of the wellbeing literature, suggesting that living a life of wellbeing, although it depends on declarative knowledge that can be codified in language, also has a high procedural component that cannot be easily codified in an expert system (Gregory, 2006).

The predictive power of datasets such as those of the ENUT depends on a broad set of activity patterns and on characteristics of the respondent's environment, where environments with greater resources are linked to greater SWB, as was glimpsed in Chapter V and as the wellbeing working definition presented in Chapter II indicates. Therefore, given this strong environmental component, it was concluded that the expert systems approach is less suitable than ML despite the challenges in implementing the latter since the intelligent time management tool being designed requires gathering procedural knowledge in the form of usage patterns of time in large datasets relevant to the domain to be evaluated and to the user's environment: contextual and specific data about the user's activities within the framework of an organization or social group are required. The relative difficulty in allowing contextual changes to an expert system makes it a less suitable option for the team's original purposes. The requirement of continuous learning and a long process in which an organization or social unit renegotiates the expectations of the different roles of its members (Aeon & Aguinis, 2017) highlights the advantages of the approach of ML. It leaves the expert system approach to much more concrete and fixed situations.

VI.2. Justification and design of target variables

The research team's previous experience with the ENUT microdata analysis and the literature presented in Chapter II of this dissertation led to a focus on the prediction of values for variables of time-use satisfaction, which are the cognitive SWB evaluations included in the training datasets that are most intimately linked to time management practices and can therefore better evaluate the performance of an intelligent time management tool. While there is some evidence that time management practices sometimes result in the preservation or enhancement of satisfaction in general or specific domains (Aeon & Aguinis, 2017; Claessens et al., 2007), how these results are obtained has been difficult to link to activities people may spend their time on (Diener, 1984) and the relationships between time use and SWB have remained more complex than anticipated (Marin & Ponce, 2020; Zuzanek & Zuzanek, 2015). On the other hand, there is empiric evidence that this linkage with time-use patterns can be done with some SWB evaluations of time-use satisfaction specifically (Marin et al., 2021; Marin & Ponce, 2020) and that this metric is then linked to other subjective and objective wellbeing evaluations (Boniwell, 2006). This link satisfies the condition presented in Chapter II for time management tools that use AI: that they have a wellbeing criterion embedded into them to, at a minimum, protect their user's wellbeing and ideally enhance it.

Given that the two environments with the most structures relative to time management are precisely work and academic environments, it is convenient to choose from these subdomains of time-use satisfaction to evaluate the performance or benefits of an intelligent time management tool. Time-use satisfaction in these environments or activities is also part of the set of SWB attributes included in the ENUT microdata.

As has already been presented, the values of these time-use satisfaction attributes in the ENUT microdata are in the form of 4 non-ordinal classes identified as "1, "2", "3," and "8". However, as already mentioned, during the preprocessing of the database the records of "8" were transformed to "0". Each class corresponds to the following situations:

0: Does not perform these types of activities

1: I would like to spend less time on these types of activities

2: The time I dedicate to this type of activity is fine

3: I would like to dedicate more time to this type of activities

At this point, it is crucial to link the design of the target values, which will be the foundation of the SWB criteria for the intelligent time management tool to be developed, to our working definition of wellbeing. As described in Chapter II, our working definition of wellbeing is:

Wellbeing is the state of human beings in which their mental and bodily conditions allow them to act while gaining a greater ability to interact with their environment.

While, in the ENUT preprocessed datasets, we already have more than one hundred thousand SWB evaluations related to time-use satisfaction in the two environments of interest, work, and academic activities, our main design principle is about agency, which is linked to overall wellbeing. So, it is essential to view the values of SWB attributes in question from the point of view of personal agency, as follows:

0: Indirectly indicates the possibility of agency in the act of refusing to carry out a specific activity. However, it may be that the environment has not presented situations where this can be tested, nor is it known whether the possibility of spending time on these activities has even been considered. It is only known that the person could refuse under current conditions because they are not doing this activity.

1: Indicates the absence of agency regarding the type of activities considered since the respondents would like to dedicate less time to them, and their environment prevents them from doing so. The respondent considers that too much time is spent on the activities in question.

2: Indicates the possibility of agency to decide the appropriate time to dedicate to certain types of activities.

3: Indicates the absence of agency regarding the type of activities considered since the respondent would like to dedicate more time to them, and her environment prevents it. The person considers the time dedicated to the activities in question insufficient.

As we can see, from a theoretical point of view, only the values “1” and “3” indicate with certainty an absence of agency, while the values “0” and “2” only indicate the possibility of the presence of agency to dedicate time, or not, to some activity. However, the value of “2” is an absolute indication of time-use satisfaction, unlike the value “0”. All other options include a clear possibility of dissatisfaction. This fact alone makes the “2” an interesting target for prediction.

According to our working definition of wellbeing, and given that Spinozist insight indicates that a greater capacity to act intentionally leads to positive affects and a lesser capacity to negative ones with their respective correlates in cognitive evaluations such as satisfaction (Diener, 1984; Diener et al., 1999), it should be expected that values for these variables that indicate a possible lack of agency must be inversely correlated to other SWB variables, while those that indicate a possible presence of agency must be directly correlated to them. During the preprocessing of the data described in Chapter V, the non-ordinal categorical variables of time-use satisfaction were divided by a process of one-hot encoding into four binary dummy variables corresponding to each of the possible values of the original categorical variable. Given these data, we have several assumptions that can be written in the form of correlations:

1. There is a positive correlation between the binary variables of time-use satisfaction linked to the possibility of agency (“0” and “2”) and the happiness variable, as well as others of satisfaction in general and in other domains.
2. The positive correlation with the happiness variable is greater for the binary variable linked to the value of “2”, backing Spinozist insight presented in Chapter II.

3. There is a negative correlation between the binary variables of time-use satisfaction linked to the absence of agency ("1" and "3") and the happiness variable, as well as others of satisfaction in general and in other domains.

In order to ascertain which of the dummy variables produced by each of the four possible values of the time-use satisfaction variables is a better target to predict wellbeing and agency through the presence of this SWB evaluation, a correlation analysis was carried out using the Pearson coefficient as recommended by the wellbeing studies literature (Ark, 2014). Regarding the correlation between SWB variables and the binary variables of satisfaction in the use of time linked to the possibility of agency ("0" and "2"), the data can be seen in **Table 19** and **Table 20**.

Table 19: *Correlations between ordinal variables of satisfaction and happiness with binary variables of satisfaction in the use of time linked to "not spending time on the group of activities"; positive correlations in bold*

Satisfaction or Happiness	Domestic activities	Study	Work	Care and support	Social	Transfers	Paperwork, Payments	Preferred activities
General	-0.036	-0.115	0.009	-0.026	-0.064	-0.067	-0.030	-0.066
Familiar	-0.031	-0.059	-0.007	-0.065	-0.104	-0.042	-0.048	-0.045
Affective	-0.030	-0.049	-0.011	-0.057	-0.094	-0.038	-0.046	-0.045
Social	-0.023	-0.110	**0.014**	**0.012**	-0.040	-0.066	-0.011	-0.064
Economic	-0.004	-0.168	**0.068**	**0.035**	-0.017	-0.052	**0.003**	-0.064
Living place	-0.010	-0.122	**0.059**	**0.023**	-0.023	-0.019	-0.019	-0.036
Happiness	-0.058	-0.127	**0.008**	-0.069	-0.098	-0.075	-0.031	-0.068

Table 20: *Correlations between ordinal variables of satisfaction and happiness with binary variables of satisfaction in the use of time linked to "being satisfied with the time spent"; larger correlations in bold*

Satisfaction or Happiness	Domestic activities	Study	Work	Care and support	Social	Transfers	Paperwork Payments	Preferred activities
General	0.0812	0.1033	**0.0585**	0.0365	0.0409	0.0496	0.0310	0.0571
Familiar	0.0306	0.0563	0.0324	0.0145	0.0370	0.0172	**0.0390**	0.0236
Affective	0.0280	0.0471	0.0356	0.0135	0.0382	0.0164	0.0356	0.0247
Social	0.0891	0.0986	0.0521	0.0306	**0.0687**	0.0639	0.0192	0.0744
Economic	**0.0997**	**0.1497**	0.0428	**0.0495**	0.0474	**0.0726**	0.0213	**0.0829**
Living place	0.0739	0.1066	0.0226	0.0392	0.0308	0.0384	0.0280	0.0580
Happiness	0.0316	0.1046	0.0323	0.0157	0.0336	0.0485	0.0302	0.0199

As can be seen in the correlations in **Table 19**, most of the correlations are negative and have very low absolute values, only not performing academic activities has negative correlations worth exploring. On the other hand, all the correlations in **Table 20** are positive, as we would expect, while not losing sight of Spinozist theory. However, most of them also have a low absolute value, so it is evidence that being satisfied with time-use patterns has some correlation with satisfaction in other domains and perceived happiness. This correlation is positive, even if it is small.

The mix of positive and negative small correlations in **Table 19** makes using the corresponding value of "Not performing these actions" a somewhat inadequate target to predict agency and wellbeing. On the other hand, the all-positive correlations in **Table 20**, with relatively large values in the area of activities related to study and work activities, make it already a much better target variable, particularly for academic activities.

The correlations between SWB variables and the binary variables of time-use satisfaction linked to the absence of agency ("1" and "3") can be seen in **Table 21** and

Table 22.

Table 21: *Correlations between ordinal variables of satisfaction and happiness with binary variables of time-use satisfaction linked to "wanting to spend less time on the group of activities" (The time spent is "too much"); positive correlations in bold*

Satisfaction or Happiness	Domestic activities	Study	Work	Care and support	Social	Transfers	Paper-work, Pay.	Preferred activities
General	-0.066	0.012	-0.063	-0.033	-0.013	**0.015**	**0.008**	-0.022
Familiar	-0.046	-0.004	-0.019	-0.048	-0.007	**0.027**	**0.025**	-0.013
Affective	-0.045	-0.002	-0.020	-0.039	-0.015	**0.022**	**0.026**	-0.021
Social	-0.070	**0.015**	-0.077	-0.020	-0.009	-0.005	-0.008	-0.015
Economic	-0.070	**0.037**	-0.088	-0.027	-0.014	-0.027	-0.033	-0.012
Living place	-0.055	**0.028**	-0.073	-0.032	-0.009	-0.025	-0.008	-0.008
Happiness	-0.056	**0.023**	-0.032	-0.038	-0.011	**0.023**	**0.010**	-0.013

Table 22: *Correlations between ordinal variables of satisfaction and happiness with binary variables of time-use satisfaction linked to "wanting to spend more time on the group of activities" (The time spent is "very little"); positive correlations in bold*

Satisfaction or Happiness	Domestic activities	Study	Work	Care and support	Social	Transfers	Paper-work, Pay.	Preferred activities
General	-0.027	**0.044**	-0.035	-0.001	**0.024**	**0.001**	-0.005	-0.021
Familiar	**0.013**	**0.026**	-0.020	**0.060**	**0.064**	-0.003	-0.001	**0.000**
Affective	**0.015**	**0.019**	-0.015	**0.051**	**0.056**	0.001	**0.001**	**0.001**
Social	-0.042	**0.040**	-0.016	-0.035	-0.026	**0.006**	-0.002	-0.040
Economic	-0.064	**0.053**	-0.082	-0.073	-0.026	-0.001	-0.006	-0.049
Living place	-0.042	**0.042**	-0.049	-0.051	-0.005	**0.003**	**0.002**	-0.039
Happiness	**0.034**	**0.057**	-0.028	**0.060**	**0.063**	**0.007**	-0.003	**0.014**

Unlike the data with binary variables of time-use satisfaction linked to the possibility of agency, the data in **Table 21** and

Table 22 show that those linked to the absence of agency in terms of the use of time have a more mixed balance of positive and negative correlations, with low absolute values without exceptions. These findings are evidence that the link between not being satisfied with the use of time and wanting to spend more or less time on a particular group of activities does not necessarily indicate a decrease in general satisfaction in particular domains or perceived happiness across the board. The study and work activities columns in

Table 22 serve as an example, considering that "very little time" is spent studying and the desire to spend more time is positively correlated with general satisfaction in all the domains whose data were collected, as well as with perceived happiness. On the other hand, the same situation of spending "very little time" at work and wanting to spend more time is negatively correlated with these same indicators of satisfaction and perceived happiness. In this case, the difference does not lie in whether one is satisfied with the use of time but in what type of activities said satisfaction is evaluated. For some, dissatisfaction and wanting a change positively correlate with other SWB indicators. Moreover, for others, dissatisfaction and wanting a change are negatively correlated with these SWB measures.

Therefore, regarding the possibilities presented above, it can be said that:

1. There is only evidence of a constant positive correlation in the binary variable linked to being satisfied with the time dedicated to the set of activities included in the ENUT survey with other SWB variables, while not devoting time to these activities has positive and negative correlations in different cases.
2. The correlations in absolute value with the happiness variable are is greater for the binary variable linked to the value of “2” (“being satisfied with the time dedicated to the set of activities included in the ENUT survey”) than for that of “0” (“Not spending time in these activities”)
3. There is a mixture of positive and negative correlations concerning the binary variables of satisfaction in the use of time linked to the absence of agency (“1” and “3”) and the happiness variable, as well as to others of satisfaction in other domains.

In most cases, the absolute values of these correlations tend to be low, which does not necessarily imply low predictive power. However, it does mean that other factors must be taken into account. This situation does not mean that these attributes are not useful for classification tasks into ordinal categories of SWB or perceived happiness, even though other attributes have a much higher correlation because they explain large amounts of variability, such as income or the presence of basic services in the home whose presence is generally the difference between the lowest levels of SWB and the average (Marin & Ponce, 2020). In this case, the marginal difference caused by the time-use satisfaction attributes may be part of the difference between rising from the average to the higher levels.

Therefore, using a binary target variable indicating the absence or presence of time-use satisfaction is warranted. In practice, this means using the ML methods to predict the binary dummy variable linked to the original value of “2”, that is, when the respondents indicate that they are satisfied with the time assigned to the activity in question, and according to the data shown in **Table 20** has a relatively significant correlation to SWB levels in general in the case of time-use satisfaction in academic activities, which is the variable that is of interest for the experiment described in this text. However, the correlations for work-related activities are also promising, even if lower in absolute values. It should be noted that while this dummy variable linked to being satisfied with the time assigned to the activities in question indicates the possibility of agency, by its very nature, it indicates the absolute presence of satisfaction, with all its correlates in SWB and objective wellbeing.

As for the predictive power of the ENUT preprocessed datasets for these and other SWB attributes, a predictive power analysis is in order.

VI.3. Predictive power analysis of training datasets

A predictive analysis of all the attributes of subjective wellbeing contained in the ENUT-2019 was carried out using the classifiers based on the intelligent algorithms seen in Chapter IV. Throughout this dissertation all the algorithms, including classifiers, feature selection and performance metrics, were implemented using Jupyter Notebooks 6.4.8 in an Anaconda environment in a Windows PC, using a Python 3.9.12 kernel, Numpy 1.21.5, and Pandas 1.4.2. All classifiers and performance metrics were the standard implementation from scikit-learn 1.0.2 and the default parameters were used unless otherwise noted in the following list:

- **Random Forest (RF):** RandomForestClassifier function. 1500 trees, a maximum of 100 levels; Gini criterion.
- **Decision tree (DT):** DecisionTreeClassifier function. Default optimized version of the CART algorithm.
- **Support vector machines (SVM):** svm.SVC function. Radial basis function (RBF) kernel, with a decision function one v.s. one shaped.
- **Deep neural networks (DNN):** MLPClassifier function. Feedforward architecture with 1000 hidden layers with as many artificial neurons in their input and hidden layers as attributes in the preprocessed dataset and a one element output layer. An L_2 penalty of 10^{-5}, a rectified linear unit function (RELU) and an adam stochastic gradient-based optimizer shuffling samples in each iteration were used.
- **Multinomial logistic regression (MLR):** LogisticRegression function. Multinomial, with a Broyden–Fletcher–Goldfarb–Shanno (LBFGS) optimization algorithm and an L_2 penalty term.

As seen in Chapter V, the ENUT-2019 data set is unbalanced regarding the attributes of general satisfaction and in particular domains –except for the time-use domain –and perceived happiness, with most of the records grouped into 1 or 2 classes. If left as is, this imbalance causes the classifiers to have deceptively satisfactory accuracy metrics, but only because they assign the vast majority of test points to the majority classes. Therefore, for all the prediction tasks whose results are presented in this work, an undersampling process was first carried out in which records from the majority classes are randomly discarded until there are classes with the same number of members or as close as possible. Since the ENUT-2019 dataset is in the tens of thousands of records, it is possible to do this for most attributes without losing much predictive power, as the high accuracy metrics seen in this section attest. All predictive analyses were performed under a ten folds cross-validation regime. Predictive analyses referring to time-use satisfaction and then all the other SWB metrics of the ENUT-2019 will be seen separately.

I.3.1. Predictive analysis of time-use satisfaction

In the case of the time-use satisfaction attributes analyzed in this section, the domain is the use of time in different groups of activities. In **Table 23**, we can see the predictive analysis for the presence or absence of time-use satisfaction in all activities for which there is a question in the ENUT questionnaire. From this point on, all computation time metrics mentioned and shown in tables refer to the test computer with an Intel Core i5-9300H processor at 2.40 GHz and 32GB of RAM. Parallel computing was used when possible, and all computations were done in a Jupiter Notebook environment with a Python 3.9.12 kernel.

Table 23: *Predictive analysis for time-use satisfaction using the complete ENUT-2019 preprocessed data set.*

Type of activities evaluated	Time use assessment	ML	Accuracy	Precision	F_1 Score	Comp. Time
Domestic	Satisfactory	RF	**0.6485**	**0.6526**	**0.6461**	101.9
Domestic	Satisfactory	SVM	0.5804	0.5804	0.5804	498.9
Domestic	Satisfactory	DNN	0.5874	0.5875	0.5873	915.9
Domestic	Satisfactory	MLR	0.5947	0.5947	0.5947	1.7
Academic	Satisfactory	RF	**0.9568**	**0.9601**	**0.9568**	14.4
Academic	Satisfactory	SVM	0.8983	0.899	0.8983	12.4
Academic	Satisfactory	DNN	0.9333	0.936	0.9331	151.9
Academic	Satisfactory	MLR	0.9143	0.9146	0.9142	0.5

Work	Satisfactory	RF	**0.8124**	**0.8617**	**0.8058**	54.5
Work	Satisfactory	SVM	0.7795	0.7896	0.7776	224.8
Work	Satisfactory	DNN	0.7639	0.7682	0.763	537.1
Work	Satisfactory	MLR	0.7349	0.7362	0.7345	1.4
Support to close ones	Satisfactory	RF	**0.6906**	**0.6906**	**0.6905**	56.7
Support to close ones	Satisfactory	SVM	0.5919	0.5924	0.5914	195.0
Support to close ones	Satisfactory	DNN	0.5873	0.5899	0.5843	91.8
Support to close ones	Satisfactory	MLR	0.6019	0.6019	0.6018	1.2
Social/family	Satisfactory	RF	**0.7133**	**0.7485**	**0.7027**	50.3
Social/family	Satisfactory	SVM	0.6306	0.6380	0.6256	210.0
Social/family	Satisfactory	DNN	0.6519	0.6519	0.6519	299.1
Social/family	Satisfactory	MLR	0.6431	0.6456	0.6416	1.3
Commutes	Satisfactory	RF	**0.8092**	**0.8259**	**0.8067**	73.0
Commutes	Satisfactory	SVM	0.7117	0.7366	0.7039	337.3
Commutes	Satisfactory	DNN	0.7323	0.7326	0.7322	777.9
Commutes	Satisfactory	MLR	0.6985	0.7004	0.6978	1.8
Pay. and paperwork	Satisfactory	RF	**0.9625**	**0.9647**	**0.9625**	15.7
Pay. and paperwork	Satisfactory	SVM	0.6053	0.6244	0.5895	36.2
Pay. and paperwork	Satisfactory	DNN	0.8443	0.8443	0.8443	114.3
Pay. and paperwork	Satisfactory	MLR	0.8415	0.8447	0.8411	0.5
Preferred	Satisfactory	RF	**0.5961**	**0.5962**	**0.596**	73.8
Preferred	Satisfactory	SVM	0.5826	0.5836	0.5814	361.0
Preferred	Satisfactory	DNN	0.5317	0.5322	0.53	568.4
Preferred	Satisfactory	MLR	0.5843	0.5843	0.5842	1.6
Domestic	Too much	RF	**0.6685**	**0.6704**	**0.6676**	11.6
Domestic	Too much	SVM	0.6463	0.6477	0.6455	13.9
Domestic	Too much	DNN	0.5708	0.582	0.5557	74.2
Domestic	Too much	MLR	0.6185	0.6192	0.6179	0.3
Academic	Too much	RF	**0.9003**	**0.9158**	**0.8994**	3.3
Academic	Too much	SVM	0.887	0.8884	0.8869	0.5
Academic	Too much	DNN	0.8588	0.8588	0.8588	10.0
Academic	Too much	MLR	0.8621	0.8651	0.8618	0.1
Work	Too much	RF	**0.7906**	**0.8073**	**0.7877**	19.9

Work	Too much	SVM	0.7691	0.794	0.7641	31.2
Work	Too much	DNN	0.7541	0.7589	0.7529	123.8
Work	Too much	MLR	0.7626	0.7675	0.7616	0.6
Support to close ones	Too much	RF	**0.6901**	**0.6902**	**0.6901**	2.9
Support to close ones	Too much	SVM	0.5718	0.5726	0.5704	0.4
Support to close ones	Too much	DNN	0.5268	0.5293	0.5181	9.8
Support to close ones	Too much	MLR	0.5859	0.586	0.5859	0.1
Social/family	Too much	RF	**0.6725**	**0.6796**	**0.6692**	2.9
Social/family	Too much	SVM	0.5409	0.5417	0.5389	0.4
Social/family	Too much	DNN	0.5117	0.5118	0.5105	18.1
Social/family	Too much	MLR	0.5439	0.5439	0.5438	0.1
Commutes	Too much	RF	**0.7668**	**0.775**	**0.7651**	37.8
Commutes	Too much	SVM	0.7573	0.7607	0.7566	101.9
Commutes	Too much	DNN	0.7214	0.722	0.7212	413.8
Commutes	Too much	MLR	0.749	0.749	0.749	1.1
Pay. and paperwork	Too much	RF	**0.9354**	**0.9422**	**0.9352**	7.5
Pay. and paperwork	Too much	SVM	0.6	0.6038	0.5963	8.5
Pay. and paperwork	Too much	DNN	0.7715	0.7869	0.7684	23.7
Pay. and paperwork	Too much	MLR	0.8456	0.8493	0.8452	0.2
Preferred	Too much	RF	**0.546**	**0.5461**	**0.5457**	3.0
Preferred	Too much	SVM	0.5201	0.5257	0.4926	0.4
Preferred	Too much	DNN	0.5201	0.5211	0.5146	12.2
Preferred	Too much	MLR	0.5489	0.5489	0.5487	0.1
Domestic	Very little	RF	**0.6493**	**0.6494**	**0.6493**	50.1
Domestic	Very little	SVM	0.5957	0.5965	0.5949	186.8
Domestic	Very little	DNN	0.574	0.5746	0.5731	441.4
Domestic	Very little	MLR	0.6043	0.6044	0.6043	1.2
Academic	Very little	RF	**0.9222**	**0.9327**	**0.9217**	5.4
Academic	Very little	SVM	0.8032	0.8037	0.8031	3.0
Academic	Very little	DNN	0.8508	0.8514	0.8507	76.1
Academic	Very little	MLR	0.8341	0.8341	0.8341	0.2
Work	Very little	RF	**0.7718**	**0.7858**	**0.769**	9.8
Work	Very little	SVM	0.6176	0.6248	0.6121	15.0

Work	Very little	DNN	0.6494	0.6595	0.6438	82.5
Work	Very little	MLR	0.5949	0.5949	0.5948	0.4
Support to close ones	Very little	RF	**0.7357**	**0.7376**	**0.7351**	67.2
Support to close ones	Very little	SVM	0.6618	0.6645	0.6604	315.3
Support to close ones	Very little	DNN	0.6591	0.6618	0.6576	605.5
Support to close ones	Very little	MLR	0.6594	0.6602	0.659	1.6
Social/family	Very little	RF	**0.7441**	**0.7833**	**0.735**	62.1
Social/family	Very little	SVM	0.6505	0.6549	0.6481	318.4
Social/family	Very little	DNN	0.6766	0.6767	0.6766	491.5
Social/family	Very little	MLR	0.652	0.6529	0.6515	1.5
Commutes	Very little	RF	**0.6703**	**0.6984**	**0.6584**	3.5
Commutes	Very little	SVM	0.5993	0.6261	0.5772	0.9
Commutes	Very little	DNN	0.6339	0.6339	0.6338	10.1
Commutes	Very little	MLR	0.6248	0.6273	0.623	0.1
Pay. and paperwork	Very little	RF	**0.8824**	**0.8864**	**0.8821**	2.2
Pay. and paperwork	Very little	SVM	0.5966	0.6074	0.5877	0.1
Pay. and paperwork	Very little	DNN	0.6723	0.6779	0.6692	4.3
Pay. and paperwork	Very little	MLR	0.6471	0.6564	0.6425	0.1
Preferred	Very little	RF	**0.6051**	**0.6058**	**0.6044**	87.8
Preferred	Very little	SVM	0.5842	0.5863	0.5817	489.7
Preferred	Very little	DNN	0.5339	0.534	0.5336	1115.5
Preferred	Very little	MLR	0.5919	0.592	0.5919	1.7

As we can see in **Table 23**, the three performance measures, accuracy, precision, and F_1 Score, are similar, so we are faced with a situation in which we can compare experiments through accuracy. That is to be expected since we are feeding the ML methods as balanced a dataset as possible.

On the other hand, we can see that the best results correspond to:

- Satisfaction in the use of time in academic activities: accuracy > 0.9
- Satisfaction in the use of time in payments and procedures: accuracy > 0.9
- Satisfaction in the use of time in work activities: accuracy > 0.8
- Satisfaction in the use of time in social activities: accuracy > 0.7
- Satisfaction in the use of time in transfers: accuracy > 0.7

The other satisfaction evaluations regarding other types of activities have accuracy metrics lower than 0.7 but greater than 0.6 in at least one variable using one of the tested ML methods, except for the one referring to time-use satisfaction in preferred activities (what the respondent likes to do); which since these are binary variables is already a low metric since a random classifier would tend to have 0.5 in the accuracy metric.

In all cases, the ML method with the best performance is random forest (RF), and the one that uses less computing time is multinomial logistic regression (MLR).

Even from this vantage point, without considering the practicalities of designing a measurement instrument as small as possible but that produces a dataset with high predictive power for a variable of interest, the question arises if there is a possibility of obtaining similar performances to those shown in **Table 23** but using a smaller sub dataset from the preprocessed ENUT-2019 dataset. Therefore, the predictive analysis was redone, but using a simple attribute reduction procedure that leads to processing only the 30 attributes with the most significant predictive power over the variable according to statistical tests. This process was carried out using the data from a statistical analysis with χ^2(chi-square) test.

For the present preprocessed ENUT-2019 dataset, the tests and the χ^2, which estimate the degree of dependence between random variables, give similar results. However, the test χ^2 generates a reduced subset of attributes in which the applied predictive tools have higher performance than with the complete preprocessed ENUT-2019 dataset, so this is the test used to generate the reduced subsets of attributes in all the predictive analyses of this work, except where the use of an alternate test is indicated. The χ^2 test, although generally used for categorical data, can also be used appropriately for integer numeric data such as frequencies (scikit-learn developers, 2020) or, in this case, the number of minutes spent on activities. A comparison of the accuracy metrics and computation times using attribute reduction and the full dataset can be seen in **Table 24**.

Table 24: *Comparison of predictive analytics for time-use satisfaction using the entire dataset and using only the best 30 attributes using one χ^2 test*

Type of activities evaluated	Time use assessment	ML	Accuracy with the complete dataset	Accuracy with Top 30 attributes	Comp. time (s)	Comp. time (s)

Domestic	Satisfactory	RF	**0.6485**	**0.6197**	101.9	33.5
Domestic	Satisfactory	SVM	0.5804	0.5814	498.9	48.0
Domestic	Satisfactory	DNN	0.5874	0.6181	915.9	101.5
Domestic	Satisfactory	MLR	0.5947	0.5824	**1.7**	**0.7**
Academic	Satisfactory	RF	**0.9568**	**0.9611**	14.4	6.0
Academic	Satisfactory	SVM	0.8983	0.9014	12.4	1.3
Academic	Satisfactory	DNN	0.9333	0.9479	151.9	40.2
Academic	Satisfactory	MLR	0.9143	0.9146	**0.5**	**0.2**
Work	Satisfactory	RF	**0.8124**	**0.8156**	54.5	21.3
Work	Satisfactory	SVM	0.7795	0.7772	224.8	22.2
Work	Satisfactory	DNN	0.7639	0.7958	537.1	92.0
Work	Satisfactory	MLR	0.7349	0.7566	**1.4**	**0.7**
Support to close ones	Satisfactory	RF	**0.6906**	**0.6413**	56.7	21.7
Support to close ones	Satisfactory	SVM	0.5919	0.5998	195.0	21.5
Support to close ones	Satisfactory	DNN	0.5873	0.638	391.8	86.6
Support to close ones	Satisfactory	MLR	0.6019	0.5993	**1.2**	**0.6**
Social/family	Satisfactory	RF	**0.7133**	**0.7107**	50.3	22.1
Social/family	Satisfactory	SVM	0.6306	0.6288	210.0	22.9
Social/family	Satisfactory	DNN	0.6519	0.6969	299.1	150.8
Social/family	Satisfactory	MLR	0.6431	0.6457	**1.3**	**0.6**
Commutes	Satisfactory	RF	**0.8092**	**0.7213**	73.0	28.8
Commutes	Satisfactory	SVM	0.7117	0.7069	337.3	46.7
Commutes	Satisfactory	DNN	0.7323	0.7123	777.9	114.5
Commutes	Satisfactory	MLR	0.6985	0.6881	**1.8**	**0.6**
Pay. and paperwork	Satisfactory	RF	**0.9625**	**0.9591**	15.7	5.7
Pay. and paperwork	Satisfactory	SVM	0.6053	0.6381	36.2	3.4
Pay. and paperwork	Satisfactory	DNN	0.8443	0.9034	114.3	47.6
Pay. and paperwork	Satisfactory	MLR	0.8415	0.8836	**0.5**	**0.2**
Preferred	Good	RF	**0.5961**	0.5485	73.8	27.9
Preferred	Good	SVM	0.5826	0.5629	361.0	44.4
Preferred	Good	DNN	0.5317	**0.5708**	568.4	111.4
Preferred	Good	MLR	0.5843	0.5711	**1.6**	**0.7**
Domestic	Too much	RF	**0.6685**	0.6269	11.6	7.3

Domestic	Too much	SVM	0.6463	**0.6514**	13.9	1.5
Domestic	Too much	DNN	0.5708	0.5847	74.2	48.9
Domestic	Too much	MLR	0.6185	0.6366	**0.3**	**0.2**
Academic	Too much	RF	**0.9003**	**0.9053**	3.3	2.9
Academic	Too much	SVM	0.887	0.8688	0.5	0.1
Academic	Too much	DNN	0.8588	0.8787	10.0	5.5
Academic	Too much	MLR	0.8621	0.8688	**0.1**	**0.1**
Work	Too much	RF	**0.7906**	0.7642	19.9	11.2
Work	Too much	SVM	0.7691	**0.7717**	31.2	3.2
Work	Too much	DNN	0.7541	0.7696	123.8	55.4
Work	Too much	MLR	0.7626	0.7564	**0.6**	**0.3**
Support to close ones	Too much	RF	**0.6901**	**0.6254**	2.9	2.8
Support to close ones	Too much	SVM	0.5718	0.6197	0.4	0.0
Support to close ones	Too much	DNN	0.5268	0.5718	9.8	4.9
Support to close ones	Too much	MLR	0.5859	0.5887	**0.1**	**0.1**
Social/family	Too much	RF	**0.6725**	**0.655**	2.9	2.6
Social/family	Too much	SVM	0.5409	0.5731	0.4	0.1
Social/family	Too much	DNN	0.5117	0.5614	18.1	3.2
Social/family	Too much	MLR	0.5439	0.4942	**0.1**	**0.0**
Commutes	Too much	RF	**0.7668**	**0.7617**	37.8	16.0
Commutes	Too much	SVM	0.7573	0.7387	101.9	10.5
Commutes	Too much	DNN	0.7214	0.7504	413.8	116.1
Commutes	Too much	MLR	0.749	0.7419	**1.1**	**0.4**
Pay. and paperwork	Too much	RF	**0.9354**	**0.9297**	7.5	4.0
Pay. and paperwork	Too much	SVM	0.6000	0.6665	8.5	0.8
Pay. and paperwork	Too much	DNN	0.7715	0.8051	23.7	54.6
Pay. and paperwork	Too much	MLR	0.8456	0.8595	**0.2**	**0.1**
Preferred	Too much	RF	0.5460	0.5402	3.0	2.7
Preferred	Too much	SVM	0.5201	0.5345	0.4	0.1
Preferred	Too much	DNN	0.5201	**0.5862**	12.2	2.7
Preferred	Too much	MLR	**0.5489**	0.569	**0.1**	**0.1**
Domestic	Very little	RF	**0.6493**	**0.6406**	50.1	22.4
Domestic	Very little	SVM	0.5957	0.5971	186.8	21.0

Domestic	Very little	DNN	0.5740	0.6087	441.4	396.7
Domestic	Very little	MLR	0.6043	0.6045	**1.2**	**0.6**
Academic	Very little	RF	**0.9222**	**0.9206**	5.4	3.8
Academic	Very little	SVM	0.8032	0.8675	3.0	0.4
Academic	Very little	DNN	0.8508	0.8103	76.1	7.6
Academic	Very little	MLR	0.8341	0.8421	**0.2**	**0.1**
Work	Very little	RF	**0.7718**	**0.7239**	9.8	5.8
Work	Very little	SVM	0.6176	0.6025	15.0	1.6
Work	Very little	DNN	0.6494	0.6874	82.5	28.1
Work	Very little	MLR	0.5949	0.593	**0.4**	**0.2**
Support to close ones	Very little	RF	**0.7357**	**0.7085**	67.2	26.8
Support to close ones	Very little	SVM	0.6618	0.6537	315.3	37.1
Support to close ones	Very little	DNN	0.6591	0.6756	605.5	509.7
Support to close ones	Very little	MLR	0.6594	0.6389	**1.6**	**0.6**
Social/family	Very little	RF	**0.7441**	**0.7433**	62.1	23.5
Social/family	Very little	SVM	0.6505	0.6470	318.4	33.9
Social/family	Very little	DNN	0.6766	0.6736	491.5	586.9
Social/family	Very little	MLR	0.6520	0.6654	**1.5**	**0.7**
Commutes	Very little	RF	**0.6703**	**0.6321**	3.5	3.3
Commutes	Very little	SVM	0.5993	0.5902	0.9	0.1
Commutes	Very little	DNN	0.6339	0.6047	10.1	5.9
Commutes	Very little	MLR	0.6248	0.5993	**0.1**	**0.1**
Pay. and paperwork	Very little	RF	**0.8824**	**0.8908**	2.2	2.4
Pay. and paperwork	Very little	SVM	0.5966	0.5546	0.1	0.0
Pay. and paperwork	Very little	DNN	0.6723	0.6975	4.3	3.2
Pay. and paperwork	Very little	MLR	0.6471	0.8151	0.1	0.0
Preferred	Very little	RF	**0.6051**	0.5748	87.8	38.6
Preferred	Very little	SVM	0.5842	0.5743	489.7	62.3
Preferred	Very little	DNN	0.5339	0.5723	1115.5	110.2
Preferred	Very little	MLR	0.5919	**0.5793**	**1.7**	**0.9**

As can be seen in the data in **Table 24** compared to those of **Table 23**, the changes in performance are minimal, particularly in the variables whose metrics were already high, such as academic, work, and paperwork activities. Also, computing time is dramatically reduced. Regarding the performance and computation time of the ML algorithms, the good results of RF persist, being the one with the best performance and MLR being the one with the shortest computation time. These results imply that it is possible to have similar performance with all ML algorithms tried while using only a small subset of attributes from those included in the preprocessed ENUT-2019 dataset. Nonetheless, the attribute reduction procedure should be performed carefully to avoid overfitting issues. This process will be described in detail later in this chapter.

Regarding time-use satisfaction prediction, three questions are up in the air.

1. If it is relevant to include in the classification task the dummy variable referring to not spending time on the type of activities in question, and if it is not relevant, would it be helpful to remove the records where this dummy variable has value 1?
2. If it is useful to have a single variable of dissatisfaction with the use of time instead of two that specify whether the dissatisfaction is due to wanting to spend more or less time on the activities in question.
3. What are the advantages and disadvantages of dividing into binary variables and having these as variables to predict instead of making a classification with four values as the original variable?

Regarding the first question, the utility of predicting the variable associated with activities not being carried out is limited in the context of time management tools where activities that are carried out are recorded, not those that are not. However, suppose this variable had to be classified. In that case, we can see in **Table 25** that the machine learning algorithms used for the other binary variables have much higher performance predicting the variable linked to “not spending time on the group of activities under consideration,” particularly using RF, with accuracy metrics above 0.9 for satisfaction with time use in all types of activities, except “preferred activities.”

Table 25: *Comparison of predictive analysis for time-use satisfaction for the binary variable linked to not spending time on the activities in question, using the complete data set and using only the best 30 attributes through* χ^2 *tests*

Type of activities evaluated	Time use assessment	ML	Accuracy with the full dataset	Accuracy with only the top 30 attributes	Computation time (s) complete dataset	Computation time (s) with only the top 30

Domestic	No activity	RF	**0.9895**	**0.9812**	7.0	3.9
Domestic	No activity	SVM	0.7462	0.7363	10.6	1.1
Domestic	No activity	DNN	0.9445	0.9487	38.7	27.1
Domestic	No activity	MLR	0.9189	0.9702	**0.4**	**0.1**
Academic	No activity	RF	**0.9994**	**1**	14.1	4.9
Academic	No activity	SVM	0.9405	0.9350	25.1	3.0
Academic	No activity	DNN	0.9922	0.9965	140.6	70.1
Academic	No activity	MLR	0.9979	0.9988	**1.3**	**1.1**
Work	No activity	RF	**0.9998**	**0.9999**	14.9	8.5
Work	No activity	SVM	0.9696	0.9778	80.2	7.0
Work	No activity	DNN	0.9979	0.9996	245.8	147.5
Work	No activity	MLR	0.9959	0.9958	**2.0**	**0.9**
Support to close ones	No activity	RF	**0.9226**	**0.9017**	47.3	13.2
Support to close ones	No activity	SVM	0.6862	0.68	208.6	20.4
Support to close ones	No activity	DNN	0.7452	0.9034	344.6	166.6
Support to close ones	No activity	MLR	0.8743	0.8982	**1.4**	**0.7**
Social/family	No activity	RF	**0.9997**	**0.9997**	33.8	8.2
Social/family	No activity	SVM	0.8032	0.8544	187.2	17.5
Social/family	No activity	DNN	0.974	0.9854	146.4	107.3
Social/family	No activity	MLR	0.9887	0.9864	**1.2**	**0.5**
Commutes	No activity	RF	**0.9967**	**0.9977**	24.2	8.0
Commutes	No activity	SVM	0.9078	0.9105	74.4	8.9
Commutes	No activity	DNN	0.976	0.9511	140.3	48.7
Commutes	No activity	MLR	0.9809	0.9743	**1.4**	**0.7**
Pay. and paperwork	No activity	RF	**0.9925**	**0.9984**	24.1	6.9
Pay. and paperwork	No activity	SVM	0.6228	0.6216	85.1	9.1
Pay. and paperwork	No activity	DNN	0.9574	0.9645	188.6	93.3
Pay. and paperwork	No activity	MLR	0.8067	0.8708	**0.9**	**0.3**
Preferred	No activity	RF	**0.6903**	**0.6386**	7.2	4.5
Preferred	No activity	SVM	0.5398	0.5479	6.3	0.7
Preferred	No activity	DNN	0.5907	0.5944	28.4	23.4
Preferred	No activity	MLR	0.6121	0.6239	**0.3**	**0.1**

A correlation analysis for all the attributes not related to subjective wellbeing of the ENUT-2014 data set of the binary variables linked to "not devoting time to these activities" illuminates a little more why the prediction of these binary variables performs better than the others. **Table 26** shows this analysis for the binary variable of "not spending time on these activities" for the chaos of domestic activities.

Table 26: *Correlations with the variable "not spending time on domestic activities"*

Code	r	Question
P7_1_1_c_0	1,000	Did you not do chores at home?
P6_5_2_2	-0.333	Did you clean or pick up the inside of your home? (Order objects, make beds, sweep, mop, shake, wash the kitchen and bathroom, among others) On weekdays.
P6_4_3_2	-0.324	Did you cook, prepare, or heat food? Or drinks? On weekdays.
P7_1_1_c_2	-0.319	How do you feel about the time you spent last week on the chores you did at home?
P6_4_4_2	-0.312	Did you serve the food, pick up, wash, dry, or arrange the dishes? On weekdays.
P6_5_2_4	-0.286	Did you clean or pick up the inside of your home? (Order objects, make beds, sweep, mop, shake, wash the kitchen and bathroom, among others) On weekends.
P6_4_3_4	-0.266	Did you cook, prepare, or heat food? Or drinks? On weekends.
P6_4_4_4	-0.265	Did you serve the food, pick up, wash, dry, or arrange the dishes? On weekends.
P6_6_3_2	-0.246	Did you sort, fold, arrange, or put away the clothes? On weekdays.
P6_6_1_2	-0.238	Did you wash, hang, or dry the clothes? (If you did it with a machine, remove the operating time.) On weekdays.
P6_5_3_2	-0.237	Did you collect, separate, throw, or burn the garbage from your home? On Weekdays.
P6_5_1_2	-0.229	Did you sweep the sidewalk, driveway, or patio of your home? On weekdays.
SEX_c_1	0.214	You are male.

SEX_c_2	-0.214	You are female.
P7_1_1_c_3	-0.180	How do you feel about the time you spent last week on…the chores you did at home?
P6_6_5_2	-0.179	Did you clean, polish, or paint footwear? (Tennis, huaraches, boots, etc.). On weekdays.
P6_5_1_4	-0.175	Did you sweep the sidewalk, driveway, or patio of your home?
P6_6_1_4	-0.164	Did you wash, hang, or dry the clothes? (If you did it with a machine, remove the operating time.) On weekends.
P6_5_4_2	-0.162	Did you care for or water pots and plants in your yard or garden? On weekdays.
P6_5_3_4	-0.161	Did you separate, throw, or burn the garbage from your home? On weekends.
P6_8_2_2	-0.156	Did you look for or do the shopping for groceries, groceries, stationery, medicines or cleaning supplies? On weekdays.
P5_8_c_7	0.154	You are unemployed and did not look for a job because you have a disability.
P6_6_3_4	-0.150	Did you sort, fold, arrange, or put away the clothes? On weekdays.
P6_6_2_2	-0.147	Did you iron the clothes? On weekdays.
P6_5_5_2	-0.146	Did you clean, feed or care for the pet(s) (companion animals) in your household? On weekdays.
P3_11_3_c_0	-0.143	Last week, did one or more people in this household need the care of another person due to a disability, chronic or temporary illness?
P6_10_5_2	-0.142	Did you close doors and windows, put padlocks, or take other measures to protect your property and home? (Secured the car, turned on the alarm) On weekdays.
P6_5_5_4	-0.132	Did you clean, feed or care for the pet(s) (companion animals) in your household? On weekends.
STOP_c_1	0.131	Are you the head of the household?

P6_10_5_4	-0.128	Did you close doors and windows, put padlocks, or take other measures to protect your property and home? (Secured the car, turned on the alarm) On weekends.

As we can see, the vast majority of the 30 attributes with the highest Pearson correlation index are precisely domestic activities, and their correlation indices are negative, except for the attribute linked to the question of time-use satisfaction in domestic activities, which correlates perfectly. Therefore, it is unsurprising that it is easy to predict this dummy variable tied to no time spent on a particular set of activities when the time use values tied to those activities are close to or equal to zero. In other words, there are strong reasons to argue that the information contained in this variable is already contained in an easily extractable form in the time-use data set for a particular type of activity.

Likewise, the only binary variable linked to not spending time on activities with an accuracy metric of less than 0.9 is that of “preferred activities” or those that the respondent “really likes to do,” which are precisely activities on which the survey does not have specific data since there are no questions about preferred activities or a variable that indicates those that are present in the survey as preferred activities. This situation is why it is not possible to link the absence of time dedicated to some activities to the activation of this type of binary variable linked to not devoting time to preferred activities for which, precisely, there is no data.

If these variables are not so useful, the question remains whether it is worth removing the records where this binary variable is active to improve computation times. **Table 27** shows the performance in this situation compared to the case in which these records are kept. In both cases, the complete data set is used.

Table 27: *Comparison of predictive analytics using the complete dataset but not considering records where the variable linked to not dedicating time is activated.*

Type of activities evaluated	Time use assessment	ML	Accuracy with non-use variable registers	Accuracy without non-use variable registers	Comp. time (s) with non-use records.	Comp. time (s) without non-use records

Domestic	Satisfactory	RF	**0.6485**	**0.6472**	101.9	79.7
Domestic	Satisfactory	SVM	0.5804	0.577	498.9	461.1
Domestic	Satisfactory	DNN	0.5874	0.5819	915.9	962.1
Domestic	Satisfactory	MLR	0.5947	0.578	**1.7**	**1.7**
Academic	Satisfactory	RF	**0.9568**	**0.9556**	14.4	14.6
Academic	Satisfactory	SVM	0.8983	0.8971	12.4	12.9
Academic	Satisfactory	DNN	0.9333	0.9354	151.9	91.5
Academic	Satisfactory	MLR	0.9143	0.9103	**0.5**	**0.6**
Of work	Satisfactory	RF	**0.8124**	**0.817**	54.5	57.2
Of work	Satisfactory	SVM	0.7795	0.7756	224.8	222.7
Of work	Satisfactory	DNN	0.7639	0.7726	537.1	549.2
Of work	Satisfactory	MLR	0.7349	0.742	**1.4**	**1.3**
Support to close ones	Satisfactory	RF	**0.6906**	**0.6849**	56.7	53.3
Support to close ones	Satisfactory	SVM	0.5919	0.5866	195.0	190.8
Support to close ones	Satisfactory	DNN	0.5873	0.6058	391.8	241.6
Support to close ones	Satisfactory	MLR	0.6019	0.6065	**1.2**	**1.2**
Social/family	Satisfactory	RF	**0.7133**	**0.7232**	50.3	51.2
Social/family	Satisfactory	SVM	0.6306	0.6257	210.0	208.3
Social/family	Satisfactory	DNN	0.6519	0.6618	299.1	562.5
Social/family	Satisfactory	MLR	0.6431	0.6456	**1.3**	**1.1**
Commutes	Satisfactory	RF	**0.8092**	**0.8097**	73.0	66.4
Commutes	Satisfactory	SVM	0.7117	0.7179	337.3	334.6
Commutes	Satisfactory	DNN	0.7323	0.7317	777.9	735.2
Commutes	Satisfactory	MLR	0.6985	0.7034	**1.8**	**1.5**
Pay. and paperwork	Satisfactory	RF	**0.9625**	**0.9650**	15.7	13.8
Pay. and paperwork	Satisfactory	SVM	0.6053	0.6241	36.2	35.2
Pay. and paperwork	Satisfactory	DNN	0.8443	0.8755	114.3	166.3
Pay. and paperwork	Satisfactory	MLR	0.8415	0.8697	**0.5**	**0.6**
Preferred	Satisfactory	RF	**0.5961**	**0.5891**	73.8	75.0
Preferred	Satisfactory	SVM	0.5826	0.5758	361.0	355.9
Preferred	Satisfactory	DNN	0.5317	0.5309	568.4	569.0
Preferred	Satisfactory	MLR	0.5843	0.5803	**1.6**	**1.5**
Domestic	Too much	RF	**0.6685**	**0.6565**	11.6	12.2

Domestic	Too much	SVM	0.6463	0.6477	13.9	14.4
Domestic	Too much	DNN	0.5708	0.6032	74.2	60.2
Domestic	Too much	MLR	0.6185	0.6477	**0.3**	**0.4**
Academic	Too much	RF	**0.9003**	**0.902**	3.3	3.5
Academic	Too much	SVM	0.887	0.8538	0.5	0.5
Academic	Too much	DNN	0.8588	0.6462	10.0	18.5
Academic	Too much	MLR	0.8621	0.8704	**0.1**	**0.1**
Work	Too much	RF	**0.7906**	**0.7891**	19.9	20.3
Work	Too much	SVM	0.7691	0.7611	31.2	31.4
Work	Too much	DNN	0.7541	0.7336	123.8	72.6
Work	Too much	MLR	0.7626	0.7644	**0.6**	**0.6**
Support to close ones	Too much	RF	**0.6901**	**0.6507**	2.9	3.2
Support to close ones	Too much	SVM	0.5718	0.5887	0.4	0.4
Support to close ones	Too much	DNN	0.5268	0.5887	9.8	8.1
Support to close ones	Too much	MLR	0.5859	0.5775	**0.1**	**0.1**
Social/family	Too much	RF	**0.6725**	**0.6374**	2.9	3.1
Social/family	Too much	SVM	0.5409	0.5292	0.4	0.4
Social/family	Too much	DNN	0.5117	0.5058	18.1	10.2
Social/family	Too much	MLR	0.5439	0.5292	**0.1**	**0.1**
Commutes	Too much	RF	**0.7668**	**0.7668**	37.8	34.4
Commutes	Too much	SVM	0.7573	0.7551	101.9	98.8
Commutes	Too much	DNN	0.7214	0.7148	413.8	248.1
Commutes	Too much	MLR	0.749	0.7389	**1.1**	**1.0**
Pay. and paperwork	Too much	RF	**0.9354**	**0.9222**	7.5	7.2
Pay. and paperwork	Too much	SVM	0.6	0.6139	8.5	8.3
Pay. and paperwork	Too much	DNN	0.7715	0.7759	23.7	104.9
Pay. and paperwork	Too much	MLR	0.8456	0.843	**0.2**	**0.3**
Preferred	Too much	RF	**0.5460**	**0.5747**	3.0	3.1
Preferred	Too much	SVM	0.5201	0.5144	0.4	0.4
Preferred	Too much	DNN	0.5201	0.5115	12.2	7.0
Preferred	Too much	MLR	0.5489	0.5345	**0.1**	**0.1**
Domestic	Very little	RF	**0.6493**	**0.6400**	50.1	50.7
Domestic	Very little	SVM	0.5957	0.5935	186.8	181.9

Domestic	Very little	DNN	0.5740	0.5662	441.4	317.8
Domestic	Very little	MLR	0.6043	0.6121	**1.2**	**1.0**
Academic	Very little	RF	**0.9222**	**0.9175**	5.4	5.1
Academic	Very little	SVM	0.8032	0.8254	3.0	2.9
Academic	Very little	DNN	0.8508	0.8706	76.1	71.1
Academic	Very little	MLR	0.8341	0.8429	**0.2**	**0.2**
Work	Very little	RF	**0.7718**	**0.7699**	9.8	9.3
Work	Very little	SVM	0.6176	0.7002	15.0	13.1
Work	Very little	DNN	0.6494	0.6788	82.5	67.3
Work	Very little	MLR	0.5949	0.6001	**0.4**	**0.4**
Support to close ones	Very little	RF	**0.7357**	**0.7419**	67.2	68.7
Support to close ones	Very little	SVM	0.6618	0.6525	315.3	306.3
Support to close ones	Very little	DNN	0.6591	0.6604	605.5	485.8
Support to close ones	Very little	MLR	0.6594	0.6655	**1.6**	**1.3**
Social/family	Very little	RF	**0.7441**	**0.7497**	62.1	62.7
Social/family	Very little	SVM	0.6505	0.6501	318.4	308.8
Social/family	Very little	DNN	0.6766	0.6954	491.5	642.3
Social/family	Very little	MLR	0.652	0.6496	**1.5**	**1.5**
Commutes	Very little	RF	**0.6703**	**0.6922**	3.5	3.7
Commutes	Very little	SVM	0.5993	0.6066	0.9	1.0
Commutes	Very little	DNN	0.6339	0.5483	10.1	49.7
Commutes	Very little	MLR	0.6248	0.6248	**0.1**	**0.1**
Pay. and paperwork	Very little	RF	**0.8824**	**0.8739**	2.2	2.2
Pay. and paperwork	Very little	SVM	0.5966	0.5966	0.1	0.1
Pay. and paperwork	Very little	DNN	0.6723	0.605	4.3	4.4
Pay. and paperwork	Very little	MLR	0.6471	0.6723	**0.1**	**0.1**
Preferred	Very little	RF	**0.6051**	**0.611**	87.8	89.9
Preferred	Very little	SVM	0.5842	0.5871	489.7	481.3
Preferred	Very little	DNN	0.5339	0.5448	1115.5	981.4
Preferred	Very little	MLR	0.5919	0.586	**1.7**	**1.5**

TESIS TESIS TESIS TESIS TESIS

As we can see in **Table 27**, the results are slightly different. In some cases, the accuracy goes down; in others, it goes up. However, it tends to go down in the initially higher yields with random forest, so we can see that, although it also means more computing time, it is useful to leave such records.

Moving on to the second question. Is it better to unify the variables if the attribute being predicted is one of dissatisfaction? This situation is particularly the case when an instrument is being designed to collect new data, and it is more economical to ask about dissatisfaction with the use of time without going into detail about whether said dissatisfaction is because they would like to spend more or less time on specific activities. **Table 28** shows a predictive analysis where the variables were unified in this way and compared with the results of both variables without merging them. In all three cases, the complete data set was used.

Table 28: *Comparison of predictive analyses of the two variables linked to time-use dissatisfaction, wanting to spend more or less time, with a new synthetic variable that integrates them.*

Type of activities evaluated	ML	Accuracy for the unified variable of dissatisfaction		Accuracy for the use of variable. “too much.”		Accuracy for the use of variable. “Very little.”	
Dataset size:		**all**	**30 best**	**all**	**30 best**	**all**	**30 best**
Domestic	RF	**0.6177**	**0.6024**	0.6269	**0.6685**	**0.6406**	**0.6493**
Domestic	SVM	0.5907	0.5914	**0.6514**	0.6463	0.5971	0.5957
Domestic	DNN	0.5671	0.5927	0.5847	0.5708	0.6087	0.574
Domestic	MLR	0.5866	0.5828	0.6366	0.6185	0.6045	0.6043
Academic	RF	**0.9253**	**0.9377**	**0.9053**	**0.9003**	**0.9206**	**0.9222**
Academic	SVM	0.862	0.8942	0.8688	0.887	0.8675	0.8032
Academic	DNN	0.8749	0.9135	0.8787	0.8588	0.8103	0.8508
Academic	MLR	0.869	0.8684	0.8688	0.8621	0.8421	0.8341
Work	RF	**0.7707**	**0.7639**	0.7642	**0.7906**	**0.7239**	**0.7718**
Work	SVM	0.7421	0.7377	**0.7717**	0.7691	0.6025	0.6176
Work	DNN	0.7213	0.7421	0.7696	0.7541	0.6874	0.6494
Work	MLR	0.7121	0.7066	0.7564	0.7626	0.593	0.5949
Support to close ones	RF	**0.7366**	**0.7085**	**0.6254**	**0.6901**	**0.7085**	**0.7357**
Support to close ones	SVM	0.6556	0.6404	0.6197	0.5718	0.6537	0.6618
Support to close ones	DNN	0.6568	0.7004	0.5718	0.5268	0.6756	0.6591
Support to close ones	MLR	0.6671	0.6387	0.5887	0.5859	0.6389	0.6594

Social/family	RF	**0.7529**	**0.7433**	**0.6550**	**0.6725**	**0.7433**	**0.7441**
Social/family	SVM	0.6609	0.6619	0.5731	0.5409	0.6470	0.6505
Social/family	DNN	0.6906	0.6882	0.5614	0.5117	0.6736	0.6766
Social/family	MLR	0.6555	0.6597	0.4942	0.5439	0.6654	0.652
Commutes	RF	**0.7631**	**0.7560**	**0.7617**	**0.7668**	**0.6321**	**0.6703**
Commutes	SVM	0.7454	0.7436	0.7387	0.7573	0.5902	0.5993
Commutes	DNN	0.7194	0.7455	0.7504	0.7214	0.6047	0.6339
Commutes	MLR	0.7473	0.7414	0.7419	0.7490	0.5993	0.6248
Pay. and paperwork	RF	**0.9282**	**0.9400**	**0.9297**	**0.9354**	**0.8908**	**0.8824**
Pay. and paperwork	SVM	0.6210	0.5909	0.6665	0.6000	0.5546	0.5966
Pay. and paperwork	DNN	0.7699	0.8040	0.8051	0.7715	0.6975	0.6723
Pay. and paperwork	MLR	0.8693	0.8623	0.8595	0.8456	0.8151	0.6471
Preferred	RF	**0.6054**	0.5623	0.5402	0.5460	0.5748	0.6051
Preferred	SVM	0.5853	0.5571	0.5345	0.5201	0.5743	0.5842
Preferred	DNN	0.5344	**0.5674**	0.5862	0.5201	0.5723	0.5339
Preferred	MLR	0.5890	0.5548	**0.5690**	**0.5489**	**0.5793**	**0.5919**

As we can see in **Table 28**, the performance did not differ much. However, it tends to improve in some types of activities and to worsen in others. Both approaches can be useful in designing an instrument to collect data about dissatisfaction.

Finally, the third question asks us about the advantages of having binary variables for satisfaction in the use of time or non-ordinal categorical variables as in the original data of the ENUT-2014. For this, we can see **Table 29**, where the results of a predictive analysis of the original non-ordinal categorical variables are seen, both with the complete dataset and the sub-dataset with only the 30 best attributes.

Table 29: *Predictive analysis of the original non-ordinal categorical variables linked to satisfaction in the use of time. Each variable has four possible categories.*

Type of activities evaluated	Time use assessment	ML	Accuracy with the full dataset	Accuracy with Top 30 attributes	Comp. time (s) complete dataset	Comp. time (s) Top 30

Domestic	Complete	RF	**0.6290**	**0.5811**	24.0	10.7
Domestic	Complete	SVM	0.4409	0.4770	66.3	5.7
Domestic	Complete	DNN	0.4990	0.4704	139.4	53.7
Domestic	Complete	MLR	0.5309	0.5366	**0.8**	**0.3**
Academic	Complete	RF	**0.5855**	**0.5623**	6.7	4.8
Academic	Complete	SVM	0.4992	0.5208	4.8	0.5
Academic	Complete	DNN	0.5324	0.5116	**33.6**	**23.1**
Academic	Complete	MLR	0.5357	**0.5623**	0.3	0.1
Work	Complete	RF	**0.6580**	0.6120	22.0	9.9
Work	Complete	SVM	0.5972	0.5878	55.3	5.6
Work	Complete	DNN	0.5754	0.5531	140.2	90.5
Work	Complete	MLR	0.6191	**0.6165**	**1.0**	**0.4**
Support to close ones	Complete	RF	**0.5718**	**0.4507**	5.1	3.6
Support to close ones	Complete	SVM	0.3901	0.4014	2.2	0.3
Support to close ones	Complete	DNN	0.3634	0.4141	13.8	9.6
Support to close ones	Complete	MLR	0.4521	**0.4507**	**0.2**	**0.1**
Social/family	Complete	RF	**0.6018**	0.5227	4.9	3.2
Social/family	Complete	SVM	0.2870	0.2723	2.2	0.2
Social/family	Complete	DNN	0.4217	0.4612	18.6	11.1
Social/family	Complete	MLR	0.4919	**0.524** 2	**0.2**	**0.1**
Commutes	Complete	RF	**0.6153**	**0.5898**	6.2	4.6
Commutes	Complete	SVM	0.5333	0.5387	4.0	0.4
Commutes	Complete	DNN	0.5223	0.5387	23.7	9.5
Commutes	Complete	MLR	0.5606	0.5515	**0.2**	**0.1**
Pay. and paperwork	Complete	RF	**0.4622**	**0.5042**	2.8	2.4
Pay. and paperwork	Complete	SVM	0.2647	0.3025	0.3	0.0
Pay. and paperwork	Complete	DNN	0.3571	0.3067	7.4	2.4
Pay. and paperwork	Complete	MLR	0.3403	0.416	**0.1**	**0.1**
Preferred	Complete	RF	**0.4345**	0.2964	4.9	3.8
Preferred	Complete	SVM	0.2691	**0.2978**	2.2	0.3
Preferred	Complete	DNN	0.2734	0.2921	24.0	18.0
Preferred	Complete	MLR	0.3353	**0.2978**	**0.2**	**0.1**

As can be seen in **Table 29**, the results in accuracy metrics are much lower, taking into account the comparison of each case to a random classifier, than with the binary variables; however, a view of the classification reports where the recall metrics of the different categories reveals that the best predictions are in the variables linked to not carrying out activities, something of limited utility for the purpose of this work, and the recall metrics of the other variables tend to be lower than with the studies of classification with the corresponding binary variables. Classification reports for RF in the case of variables linked to satisfaction in the use of time in domestic and academic activities, variables in which there were moderate and high performances, respectively, can be seen in **Table 30**.

Table 30: *Classification reports and confusion matrices for satisfaction with the use of time in domestic and academic activities, using the original variable with four classes.*

Classification report - evaluation of the use of time, domestic activities , precision recall f 1-score support 0 0.9200 0.9979 0.9574 956 1 0.5365 0.5309 0.5337 955 2 0.5012 0.4550 0.4770 956 3 0.5226 0.5319 0.5272 955 accuracy 0.6290 3822 macro avg 0.6201 0.6289 0.6238 3822 weighted avg 0.6201 0.6290 0.6239 3822 Confusion Matrix [[954 0 0 2] [19 507 221 208] [31 236 435 254] [33 202 212 508]] Execution time: 23.97	Classification report - evaluation of use of time, academic activities precision recall f 1-score support 0 0.9319 1.0000 0.9647 301 1 0.4681 0.5615 0.5106 301 2 0.4167 0.3322 0.3697 301 3 0.4821 0.4485 0.4647 301 accuracy 0.5855 1204 macro avg 0.5747 0.5855 0.5774 1204 weighted avg 0.5747 0.5855 0.5774 1204 Confusion Matrix [[301 0 0 0] [3 169 64 65] [7 114 100 80] [12 78 76 135]] Execution time: 6.73

As we can see in **Table 30**, compared to the values in **Table 23** there is value in designing a data collection instrument that can be effectively analyzed in binary variables, as long as the value of the degree of satisfaction or dissatisfaction is not required. On the other hand, it remains to be asked if, in the case of a multiclass classification work on these variables, it would be worthwhile to unify the classes of dissatisfaction. This unification was done, and a predictive analysis was performed, the results of which are in **Table 31**.

Table 31: *Comparison of the predictive analyses, using the complete data set, of the variables of satisfaction with the use of time with the four original non-ordinal categories and with three categories, unifying the two categories of dissatisfaction.*

Type of activities evaluated	Time use assessment	ML	Accuracy with four categories	Accuracy with three categories	Comp. time with four categories	Comp. time with three categories
Domestic	Complete	RF	**0.6290**	**0.6355**	24.0	10.7
Domestic	Complete	SVM	0.4409	0.4477	66.3	5.7
Domestic	Complete	DNN	0.4990	0.5181	139.4	53.7
Domestic	Complete	MLR	0.5309	0.5481	**0.8**	**0.3**
Academic	Complete	RF	**0.5855**	**0.5855**	6.7	4.8
Academic	Complete	SVM	0.4992	0.5008	4.8	0.5
Academic	Complete	DNN	0.5324	0.5216	33.6	23.1
Academic	Complete	MLR	0.5357	0.559	**0.3**	**0.1**
Work	Complete	RF	**0.6580**	**0.6542**	22.0	9.9
Work	Complete	SVM	0.5972	0.5906	55.3	5.6
Work	Complete	DNN	0.5754	0.5633	140.2	90.5
Work	Complete	MLR	0.6191	0.6165	**1.0**	**0.4**
Support to close ones	Complete	RF	**0.5718**	**0.5634**	5.1	3.6
Support to close ones	Complete	SVM	0.3901	0.3803	2.2	0.3
Support to close ones	Complete	DNN	0.3634	0.4014	13.8	9.6
Support to close ones	Complete	MLR	0.4521	0.4465	**0.2**	**0.1**
Social/family	Complete	RF	**0.6018**	**0.5622**	4.9	3.2
Social/family	Complete	SVM	0.2870	0.3880	2.2	0.2
Social/family	Complete	DNN	0.4217	0.4246	18.6	11.1
Social/family	Complete	MLR	0.4919	0.4963	**0.2**	**0.1**
Commutes	Complete	RF	**0.6153**	**0.6162**	6.2	4.6
Commutes	Complete	SVM	0.5333	0.5187	4.0	0.4
Commutes	Complete	DNN	0.5223	0.4923	23.7	9.5
Commutes	Complete	MLR	0.5606	0.5479	0.2	0.1
Pay. and paperwork	Complete	RF	**0.4622**	**0.5126**	2.8	2.4
Pay. and paperwork	Complete	SVM	0.2647	0.2731	0.3	0.0
Pay. and paperwork	Complete	DNN	0.3571	0.3445	7.4	2.4
Pay. and paperwork	Complete	MLR	0.3403	0.3529	**0.1**	**0.1**
Preferred	Complete	RF	**0.4345**	**0.4259**	4.9	3.8

Preferred	Complete	SVM	0.2691	0.2993	2.2	0.3
Preferred	Complete	DNN	0.2734	0.3007	24.0	18.0
Preferred	Complete	MLR	0.3353	0.2935	**0.2**	**0.1**

The accuracy metric improves using only three categories instead of four; however, once again, it must be considered that while for four categories, a random classifier with a uniform distribution would have an average accuracy of 0.2500, for three categories, it is 0.3333. In comparison, there is an advantage to only having three categories, but it is only a little.

If we have three categories, the options are reduced to satisfaction, dissatisfaction, and not having spent time on the activities in question. As we have seen, the categories of not having spent time on the activities in question do not encode information that cannot be easily inferred by seeing whether those activities' time use variables are zero, so we can justifiably withdraw them from the classification task. This situation leads us to stay with only two categories: satisfaction and dissatisfaction, which can be seen as the prediction of a binary categorical variable. The results of the predictive analysis for these binary variables are found in **Table 23** and **Table 24**, from where we had already concluded that if this approach was the best option, the best results were for satisfaction in the use of time in payments and paperwork, work, social, and academic activities.

In **Table 32**, we can see the data from **Table 24** corresponding to the use of the RF method, which in all cases showed the best performance, and with the binary variables for each group of activities in consecutive rows.

Table 32: *Performance comparison of time-use satisfaction binary variables using RF with both the complete dataset and the top 30 attributes.*

Type of activities evaluated	Time use assessment	ML	Accuracy with the full dataset	Accuracy with Top 30 attributes	Comp. time (s)	Comp. time (s)
Domestic	Satisfactory	RF	0.6485	0.6197	101.9	33.5
Domestic	Too much	RF	**0.6685**	0.6269	**11.6**	**7.3**
Domestic	Very little	RF	0.6493	**0.6406**	50.1	22.4
Academic	Satisfactory	RF	**0.9568**	**0.9611**	14.4	6.0
Academic	Too much	RF	0.9003	0.9053	**3.3**	**2.9**

Academic	Very little	RF	0.9222	0.9206	5.4	3.8
Work	Satisfactory	RF	**0.8124**	**0.8156**	54.5	21.3
Work	Too much	RF	0.7906	0.7642	19.9	11.2
Work	Very little	RF	0.7718	0.7239	9.8	5.8
Support to close ones	Satisfactory	RF	0.6906	0.6413	56.7	21.7
Support to close ones	Too much	RF	0.6901	0.6254	**2.9**	**2.8**
Support to close ones	Very little	RF	**0.7357**	**0.7085**	67.2	26.8
Social/family	Satisfactory	RF	**0.7133**	**0.7107**	50.3	22.1
Social/family	Too much	RF	0.6725	0.655	2.9	2.6
Social/family	Too much	RF	0.6725	0.655	**2.9**	**2.6**
Commutes	Satisfactory	RF	**0.8092**	0.7213	73.0	28.8
Commutes	Too much	RF	0.7668	**0.7617**	37.8	16.0
Commutes	Very little	RF	0.6703	0.6321	**3.5**	**3.3**
Pay. and paperwork	Satisfactory	RF	**0.9625**	**0.9591**	15.7	5.7
Pay. and paperwork	Too much	RF	0.9354	0.9297	7.5	4.0
Pay. and paperwork	Very little	RF	0.8824	0.8908	**2.2**	**2.4**
Preferred	Satisfactory	RF	0.5961	0.5485	73.8	27.9
Preferred	Too much	RF	0.546	0.5402	**3.0**	**2.7**
Preferred	Very little	RF	**0.6051**	**0.5748**	87.8	38.6

From **Table 32**, we can infer the following results:

For the binary variables linked to satisfaction in the use of time in the types of activities with the highest performance (academic, work, and payments and paperwork), the best performance occurs when predicting the binary variable linked to being satisfied with the use of time in these activities using both the complete dataset and a reduction of attributes to the 30 most predictive using data from a statistical analysis with χ^2 tests. On the other hand, this is true for another group of activities that present high performance in the classification task, which are commutes but only using the complete data set.

On the other hand, the binary variables linked to time-use satisfaction in activities that registered lower performance in the classification task have their best performance in the binary variables linked to dissatisfaction, either because they perceive that too little or too much time is spent and want to dedicate more or less time respectively. These results are summarized in **Table 33**.

Table 33: *Comparison of the binary variables linked to satisfaction in the use of time with the best performance in each group of activities.*

Type of activities evaluated	ML	Best Action with the full dataset	Best evaluation	Best evaluation	Best Action with 30 best attributes	The largest difference to measure var. "Good"
Domestic	RF	0.6685	Too much	0.6406	Very little	0.0209
Academic	RF	**0.9568**	Satisfactory	**0.9611**	Satisfactory	0
Work	RF	0.8124	Satisfactory	0.8156	Satisfactory	0
Support to close ones	RF	0.7357	Very little	0.7085	Very little	0.0672
Social/family	RF	0.7133	Satisfactory	0.7107	Satisfactory	0
Commutes	RF	0.8092	Satisfactory	0.7617	Too much	0.0404
Pay. and paperwork	RF	0.9625	Satisfactory	0.9591	Satisfactory	0
Preferred	RF	0.6051	Very little	0.5748	Very little	0.0263

As shown in **Table 33**, the best performance occurs when predicting the binary variable linked to considering the use of time for a particular activity group as satisfactory. Given that, as previously seen, merging the dissatisfaction variables does not entail a significant disadvantage in terms of predictive performance, it can be concluded that for the design of an instrument that collects from its respondents whether or not they are satisfied with the use of the time, the most efficient would be a yes/no question, instead of several categories that explain the reason for the dissatisfaction as we can see in **Table 34** where predicting a unified variable of dissatisfaction has lower performance than predicting the binary variable of satisfaction, which in this case is equivalent to the dummy variable linked to presence of time-use satisfaction.

Table 34: *Comparison of accuracy metrics for the best performances in terms of original binary variables and the synthetic dissatisfaction variable.*

Type of activities evaluated	ML	dataset, prediction of satisfaction or dissatisfaction	Reduced dataset, prediction of satisfaction or dissatisfaction	Dissatisfaction prediction (unified variable)

Dataset size:		Accuracy	Better ev.	Accuracy	Better ev.	Complete dataset	Reduced dataset
Domestic	RF	0.6685	Too much	0.6406	Very little	0.6177	0.6024
Academic	RF	0.9568	Satisfactory	**0.9611**	Satisfactory	0.9253	0.9377
Of work	RF	0.8124	Satisfactory	0.8156	Satisfactory	0.7707	0.7639
Support to close ones	RF	0.7357	Very little	0.7085	Very little	0.7366	0.7085
Social/family	RF	0.7133	Satisfactory	0.7107	Satisfactory	0.7529	0.7433
Transfers	RF	0.8092	Satisfactory	0.7617	Too much	0.7631	0.756
Payments and procedures	RF	**0.9625**	Satisfactory	0.9591	Satisfactory	**0.9282**	0.94

VI.3.2. Analysis of the attributes corresponding to satisfaction not related to the use of time

As has been seen, the ENUT-2014 contains attributes of SWB linked to satisfaction in general and by domains, one of which is the use of time. So far, good results have been seen in the satisfaction attributes in the time use domain, and two probable reasons have been elucidated for which predictive analyses of these attributes have high performance: they can be analyzed as binary variables, and the data set of the ENUT-2014 contains, for the most part, time use attributes.

However, when analyzing the personal satisfaction data in general and from the other domains included in the ENUT-2014, such as family life, affective life, social life, economic situation, and housing, we find low performance in the classification task, as seen in **Table 35**.

Table 35: *Predictive analysis of satisfaction in different life domains, other than time use, and overall perceived happiness using the complete and reduced data sets. All attributes have five ordinal classes.*

SWB component	Domain	ML	Accuracy with the complete dataset	Accuracy with Top 30 attributes	Comp. time (s) complete dataset	Comp. time (s) Top 30 attributes
Satisfaction	Life in general	RF	**0.3465**	**0.2618**	4.0	3.7
Satisfaction	Life in general	SVM	0.2335	0.2185	1.3	0.2

Satisfaction	Life in general	DNN	0.2467	0.226	8.6	13.5
Satisfaction	Life in general	MLR	0.2429	0.2505	0.2	0.1
Satisfaction	Family life	RF	**0.3164**	0.275	4.0	3.7
Satisfaction	Family life	SVM	0.2787	**0.2618**	1.3	0.2
Satisfaction	Family life	DNN	0.2411	0.1902	13.0	5.5
Satisfaction	Family life	MLR	0.2524	0.307	0.1	0.1
Satisfaction	Affective life	RF	**0.2938**	**0.2486**	4.1	3.7
Satisfaction	Affective life	SVM	0.2373	0.2429	1.3	0.2
Satisfaction	Affective life	DNN	0.2316	0.2373	14.8	7.1
Satisfaction	Affective life	MLR	0.2618	0.2373	0.2	0.1
Satisfaction	Social life	RF	**0.3164**	**0.2429**	4.0	3.8
Satisfaction	Social life	SVM	0.275	0.2279	1.3	0.2
Satisfaction	Social life	DNN	0.2392	0.2128	26.0	5.2
Satisfaction	Social life	MLR	0.2316	0.2298	0.1	0.1
Satisfaction	Economic situation	RF	**0.2806**	**0.2147**	4.0	3.7
Satisfaction	Economic situation	SVM	0.2618	0.2335	1.3	0.2
Satisfaction	Economic situation	DNN	0.2392	0.2203	11.9	11.4
Satisfaction	Economic situation	MLR	0.2429	0.2806	0.1	0.1
Satisfaction	Living place	RF	**0.2881**	**0.2354**	4.0	3.8
Satisfaction	Living place	SVM	0.1808	0.2542	1.3	0.2
Satisfaction	Living place	DNN	0.2316	0.2279	21.2	19.8
Satisfaction	Living place	MLR	0.2731	0.2674	0.2	0.1
Happiness	General	RF	**0.3258**	0.2576	2.9	3.0
Happiness	General	SVM	0.2348	0.2121	0.3	0.0
Happiness	General	DNN	0.2311	0.2462	5.9	3.3
Happiness	General	MLR	0.2273	**0.2803**	0.1	0.1

In contrast to their time-use satisfaction counterparts, the attributes in **Table 35** contain five ordinal classes, so the value of separating them into binary classes would be limited. In addition, the ENUT questionaries do not contain many questions directly related to the domains in which the respondent is asked to assess their satisfaction: for instance, there are no questions about how they lead their family or affective life, nor about the quantity and quality of social and family relationships. Some questions referring to economic and housing situation are included, as well as some conditions and circumstances in the home; however, not in the large number in which they are included for the time-use domain because the ENUT is a survey focused precisely on the use of time domain and secondarily on the others. Therefore, it is unsurprising that the accuracy metric is much lower than for time-use satisfaction attributes, hovering around 0.3 when a random classifier would have an average accuracy metric of 0.2.

VI.3.3. Model analysis with reduced dataset considerations

Since, in all the different predictive analyses of the previous sections, in all cases, the performance using the reduced data set is very similar to that of using the complete dataset, it is appealing to conclude that it is possible to design an instrument to capture data from new respondents based only on the most predictive attributes according to the attribute reduction process with statistical analysis with f or χ^2 tests that estimate the degree of dependence between random variables. For example, in **Table 36**, we can see the 30 attributes judged most predictive by the f test for the classification task corresponding to the attribute of satisfaction with the use of time in academic activities.

Table 36: *The 30 attributes with the greatest predictive power according to statistical tests with the f test for the classification task related to satisfaction in the use of time in academic activities.*

f Score	Questions
38855.81	Did you study or take courses or classes? (include open or distance system, diploma, etc.)*
36515.37	Are you currently attending school?
23797.04	Did you spend the last week studying? [Instead of working]
15913.69	Did you commute to school?
14838.45	Did you do homework, school practices, or any other study activity?
10434.78	Currently, are you this single?
7620.393	How old are you?
7253.877	Are you the head of the household?
5547.982	Did you do homework, school practices, or other study activities from Monday to Friday?
3109.446	Currently, are you married?
2910.961	During the past week, how much total time did you spend working?

2525.396	You do not work.
2525.396	I do not work to earn money or help with household expenses.
2525.396	So last week, were you in any other situation [study or unpaid work] not mentioned in the survey other than paid work?
2314.470	During the past week, did you work at least one hour?
2314.470	Did you work last week to earn money or help with household expenses?
2173.644	How much do you earn or receive for working?
1894.47	Did you check your email, consult social networks, or chat without doing any other activity from Monday to Friday?
1840.775	Did you check your email, social media, or chat without doing any other activity over the weekend?
1529.128	Did you study, take courses or classes at the weekend? (Include open or distance system, diploma, etc.) on the weekend?
1347.816	Did you do sports or physical exercise?
1255.12	So, did you do household chores or care for family members instead of working last week?
1243.213	Did you consult information or browse the internet without doing any other activity?
1119.471	Did you cook, prepare, or heat food or drinks Monday through Friday?
1082.798	During the past week, how much time did you spend commuting to and from work (to your activity)?
1080.427	During the past week, how much time did you spend working (your activity)?
1077.568	Did you cook, prepare, or heat food or drinks over the weekend?
887.184	Did you work as a worker or employee in your main job or activity last week?

It is tempting to use the 30 questions in **Table 36** to almost automatically generate a set of attributes to be collected by an instrument that may even have less than 30 questions, taking into account the overlaps of several of the attributes included in **Table 36** and thereby avoiding a questionnaire with more than 300 questions such as the original ENUT. However, this falls within the cases of stepwise regression and similar procedures, which in recent years have been questioned in the statistical and *Big Data* communities as procedures that could lead to misleading results, particularly when trying to use in predictive and prescriptive analysis the subset of attributes resulting from procedures such as the one used in this work to generate those in **Table 36** with a new dataset that only contains this subset of reduced attributes (Smith, 2018).

The key is that the error consists in trying to generate automatically, or almost automatically, a new regression model with a subset of attributes judged to be the most predictive utilizing statistical tests since this entails using the same data twice, first to generate the most relevant subset of attributes and then to do the regression or classification task on these attributes, which can lead to generating misleading performance metrics (Smith, 2018), and to a situation where relying on such performances an instrument is designed based solely on said subset of data and, once the time and resources have been invested in applying it to obtain a new dataset, its performance and predictive power is worse than expected.

Smith suggests other methods, first, taking domain knowledge into account to include specific attributes in a reduced dataset with the support of theory and expert knowledge. Suppose such expert knowledge is absent or hard to come by. In that case, methods such as recursive attribute elimination with cross-validation and regularized tree ensembles are also recommended, which have shown greater success even when validated with data outside the initial sample (Smith, 2018). These recommendations do not mean that in an attribute reduction process, results such as those in **Table 36** cannot be available information during the decision process involved in this attribute reduction. However, they cannot be the only factor. It must be considered that the predictive power assigned to these attributes may be an accident of the present dataset or sample and not necessarily intrinsic to the phenomena studied, so that in other datasets, even those generated by the same instrument, it is not guaranteed that the predictive power expected will reoccur. Nonetheless, applying these and other attribute reduction tools and validation procedures is necessary to arrive at an adequate instrument to generate smaller datasets with similar predictive power regarding attributes of interest.

VI.3.4. Conclusions of the predictive analysis

From the analysis of the preprocessed ENUT datasets, it is concluded that with datasets on time use and a limited number of demographic attributes, the most favorable results in terms of SWB prediction occur when predicting the presence or absence of time-use satisfaction. In contrast, the success in predicting satisfaction in other domains or perceived happiness is quite limited. Similarly, the evidence indicates that better performance is obtained by measuring and treating time-use satisfaction as a binary variable that encodes the presence or absence of time-use satisfaction in some set of activities. Suppose the instrument's design requires that satisfaction be measured in degrees due to psychological factors or for the respondents' comfort. In that case, a system analogous to that implemented at the MIT Media Lab (Sano, 2016) to transform ordinal or numeric variables into binary categorical variables can be designed considering specific percentiles assigned to one and the other value of a binary variable.

There is a relationship between a greater number of attributes related to the domain evaluated by a satisfaction judgment and a higher performance of tasks for predicting said satisfaction judgment. In this case, more than 80% of the attributes of the ENUT microdata are time-use attributes, so it is not surprising that this is precisely the domain in which predictive tools have the best performance over time. This abundance of attributes is not the case with domains of satisfaction, such as life in general, family life, affective life, and others, where predictive performance is low.

Finally, while good performances in predictive tasks were preserved by applying feature reduction employing pre-feed statistical tests, current literature (Smith, 2018) questions the appropriateness of this method, particularly when planning to build new datasets based only on these reduced attribute subsets, but also even when applying the prediction tasks with the same data set with the reduced attributes.

These results lead to the conclusion that although the attribute reduction documented in the previous section produces good results in terms of performance and less computation time, it is not possible to affirm that these results will be similar when applying the same predictive methods on a dataset from a sample external to the ENUT microdata. Methods that are less automated or incorporate validation with two different samples or cross-validation should be considered, and that is why is important that we have two different ENUT samples.

In summary:

- If a large quantity of time-use data is available (on the order of the ENUT microdata), the component of SWB that can be most effectively predicted is time-use satisfaction.
- In this case, measuring and treating time-use satisfaction as a binary variable is better.

- Feature reduction should only be used for predictions after it has been found that their performance is like that of using the full dataset, and it should not be assumed that this situation carries over to predictions with data sets that are not part of the original sample, without doing a validation procedure.

It should be stressed that there remains the possibility of carrying out a careful process of attribute reduction where the risks of using a process of automated attribute reduction are mitigated in such a way that the design of a model with fewer attributes than the preprocessed ENUT datasets can be validated and preserves a similar or greater predictive power using datasets foreign to the preprocessed ENUT datasets. This mitigation requires using techniques other than the statistical tests used in this section, such as the chi-square or the tests f, or the incorporation of expert knowledge from the domains of time use, time management, agency, and wellbeing. This very question was the topic of previous work (Marin et al., 2021) of the research team, and the strategy developed will be implemented later in this chapter. However, the design of a measurement instrument begins with setting goals and objectives, as described in the next section.

VI.4. Measurement instrument goal setting

In previous research (Marin et al., 2021) from the research team on which this section is based, it was stressed that virtually all data sets used in the context of massive data (*Big Data*) and ML are obtained through measurement instruments, which are tools that allow assigning numerals to objects or events according to specific rules (Stevens, 1946), with a greater or lower degree of methodological design. The design of measurement instruments is an essential discipline within scientific research. It is a process that cannot be solved simply without a rigorous follow-up of steps in a research methodology arbitrarily assigning numbers to phenomena (Martínez Rizo, 2010). It is generally recommended to have a study objective in mind before even beginning the first step of deciding what type of measurement instrument to design (Vogt et al., 2012), even when relevant data and indicators generated by pre-existing measurement instruments are already available (Martínez Rizo, 2010). The measurement instrument design process adheres to the process of applying the scientific method, where the formulation of hypotheses precedes the design of measurements, experiments, and predictions.

The first order of business in designing a measurement instrument is to define the objective or goal of the instrument. In other words, to define clearly what the instrument should measure. While there may be an ulterior motivation in what the instrument is measuring, the objective or goal of the instrument itself is always to obtain accurate values of what is being measured to produce useful data that can further research objectives and serve as a reference for action. While the intentions of having the measurement instrument might be greater than just obtaining the measurements (for example, gathering evidence or describing a particular situation), ideally, the instrument needs to be agnostic regarding these intentions.

Good measurement instruments produce data that can be used for purposes beyond the scope of the original intentions of the instrument designers, even as these new intentions are circumscribed by the instrument's limitations designed to fulfill its original goals. This characteristic is important because, as mentioned in previous sections of this chapter, it is rather difficult and costly to obtain original data for research purposes, particularly in exploratory phases, not just because proper data gathering is an arduous process but because of the need to justify what is going to be measured.

From a data analytics perspective, in a data-rich environment, it is fair to question whether an extensive dataset produced by a measurement instrument with excellent design can contain helpful information about what subsets of the instrument can serve for more focused intentions than those that initially led to the design of the instrument. In fact, such information is contained in such big datasets, as we will see in the next section, and in such a way that the clarification of the ulterior motivations for producing a smaller, more focused measurement instrument becomes the most critical factor as the instrument is already designed to measure specific values.

In the case of the research presented in this dissertation, this intention or motivation is to measure the values of a small subset of the attributes contained in the ENUT preprocessed datasets that can lead to accurate predictions of the presence of time-use satisfaction in work and academic environments, which were shown in the previous section as two of the best-performing classification processes with all the AI algorithms tried, in particular with the decision tree-based ones, which happen to be interpretable ML. In other words, we are engaging in model selection in obtaining a smaller measurement instrument through an analysis of the data produced by a bigger original one.

VI.5. AI-aided measurement instrument design

In the context of the work presented in this dissertation and previous research (Marin et al., 2021) on which this section of the chapter is based, AI-aided measurement instrument design is a model selection and validation process. Model selection is the process of choosing an appropriate mathematical model from a class or set of models (Sammut & Webb, 2017c). In the case addressed by this work, we will be talking about an automated model selection process. One of the tools for model selection are attribute selection strategies, which are procedures by which an increasingly common problem is addressed: many datasets of research relevance and in which ML tools could be applied have a large number of attributes or dimensions, which makes their processing difficult, decreases the performance of regression and ML tools, and increases the use of computational resources in their processing (Smith, 2018). Attribute selection aims to choose a subset of relevant attributes from among the original ones, discarding irrelevant, redundant, or noisy attributes (Wang et al., 2017).

Several attribute selection approaches can be classified according to the availability of labeled data or by the type of search strategy. Automated cases are considered in the matter that concerns this work because the labels of our variables of interest (time-use satisfaction) are available, and the prediction is a binary classification task. For this type of search strategy, we will talk about filtering tools with which the attributes are selected based on the characteristics of the data without using ML algorithms; and the wrapper ones, which use a specific ML algorithm to evaluate the quality of the selected features (Wang et al., 2017). Likewise, step-by-step regression strategies will be seen. These methods were chosen because of their popularity in the data mining community from which they have been brought to AI practice.

VI.5.1. Avoiding the risks of automated model selection

In the research team's previous engagement with these themes (Marin et al., 2021), we have remarked that large publicly available datasets or previous research datasets are often available for prediction and classification tasks assisted by AI, but advancing current research work requires collecting new data from different samples of the same population. Often, it is desirable that these new datasets have fewer attributes than the original sources. This desire leads to the question of choosing a subset of attributes that retains the predictive power of the original data set. This process is known as the model selection problem in data mining, and it appears to be a problem for automated feature selection methods.

In the context of data mining, the literature proposes solutions such as stepwise regression or attribute selection methods by filtering (Cios et al., 2007a). However, there is also consistent criticism in the literature of the use of this type of technique, even for the prediction of data within a subset of the same sample (Harrel, 1996; Smith, 2018) which warns of dangers such as overfitting and a tendency to reach false conclusions. Smith recently carried out a Monte Carlo simulation where he generated datasets with between 10 and 1000 candidate explanatory variables whose values were randomly drawn from normal distributions. Five explanatory variables were randomly chosen, and an objective function was created based on these. Next, the stepwise regression attribute selection method was applied and found that even from datasets with 100 candidate explanatory variables, the probability that a selected variable is truly explanatory is nearly equal to that of it not being an explanatory variable with no real relationship to the target variable. The situation worsens as the number of candidate explanatory variables increases (Smith, 2018), contrary to what we can find in parts of the Big Data literature and the intuition of some analysts that “more data is better.” This criticism extends to the use of other more sophisticated feature selection techniques. However, Smith himself acknowledges that techniques such as regularized decision tree ensembles and cross-validated recursive feature elimination are more successful in generating smaller attribute sets that preserve predictive and descriptive power (Smith, 2018).

However, the main obstacle to using these techniques resides in statistical issues. Many attribute selection techniques suffer from the problem of having post hoc hypotheses, that is, trying to test the hypothesis with the same data set that suggests it (Vogt, 2005). This problem is evident in stepwise or filtered regression techniques, where a subset of attributes is selected based on statistical tests with the same sample with which it is later intended to make predictions. This use of univariate tests also entails violating several of their assumptions with their use together, in addition to the fact that they assume pre-specified hypotheses. Likewise, applying these techniques can be a pretext to avoid thinking about the model selection problem as such (Harrel, 1996); exactly the opposite is needed in analysis for situations that deal with complex human realities such as SWB and its strong environmental component.

However, it can be argued that these problems are less relevant in the case of prediction with new data, although depending on the ultimate objectives of the research project, the use of automated attribute selection strategies could introduce restrictions on the type of conclusions that can be reached. By themselves, automatic attribute selection techniques are insufficient to ensure that their suggested model leads to an appropriate measurement instrument to collect new data that retains predictive power regarding the variable of interest. Something that the researcher would only learn when evaluating the performance of the new dataset. However, as long as the false confidence of just wanting to "justify with the numbers" the new model is avoided, it is possible to rely on automated attribute selection tools, which brings the advantages associated with automating part of the process of measuring instrument design in theoretical research in the application domain, if some conditions are taken into account, as we will see on the following section.

VI.5.2. Implementation of model selection

As already seen in previous research (Marin et al., 2021), there is a strong argument that the actual "gold standard" for feature selection is expert knowledge in the application domain (Smith, 2018). This affirmation is in line with the literature on the design of measurement instruments, where the operationalization of concepts is a process that must be carried out judiciously, justifying with prior evidence and theory the inclusion or exclusion of each variable that operationalizes a concept that is intended to be measured (Martínez Rizo, 2010; Roskam, 1989; Smith, 2018; Vogt, 2005)

However, automated model selection strategies have already been studied for these same purposes and are considered valid if certain statistical conditions are met. Furthermore, there is evidence that including specific attributes may not be derivable from theory and expert knowledge and require empirical evidence (Castle et al., 2011), that is, numerical and statistical analyses. Likewise, many criticisms exist of how concepts are commonly operationalized in the Social Sciences. However, it is recognized that despite all the methodological and epistemological obstacles, in the end, the data can have correlations with real behaviors and predictive validity even if the fidelity of the data continues to be a topic of epistemological or investigative discussion: Ultimately, a theory is more than an inductive generalization, rather it must describe, explain and predict; and there being no profound difference between these three there is value (Roskam, 1989) in a model that has practical predictive power as long as its statistical or inferential limitations are respected (Marin et al., 2021).

Attending to the steps of the measurement instrument design process and the limitations of statistical tools is what guarantees predictive power, so there are certain conditions under which automated attribute selection tools can be used for the present objective:

a) The original data set was collected utilizing a measurement instrument whose design already considers statistical theory and the necessary expert knowledge. This advantage does not exempt from studying the instrument and its theory.

b) There are at least two different samples. One to generate the subset of attributes and another to validate the performance. In the absence of two different samples, a tolerable substitute would be to separate a single sample with a large number of records into two subsets. Depending on the attribute to be predicted, it may be necessary to have several thousand or tens of thousands of records for this purpose.
c) The use of attribute selection strategies that widely violate the assumptions of the statistical tests used, even if their results appear good, should be avoided as much as possible. This includes filtering techniques by statistical tests and stepwise regression.
d) The analyst should have at least a general knowledge of the application domain and access to expert knowledge through literature sources or, ideally, in person.
e) The analyst must have a general knowledge of the subject of measurement instrument design and access to expert knowledge.
f) The necessary time must be allowed to make relevant modifications to the model suggested by the automatic attribute selection methods. (Marin et al., 2021)

Once these conditions are met, the following six steps can be followed:

1) Ensure the datasets have been adequately cleaned, preprocessed, and tested for predictive power, preferably with cross-validation regimes.
2) Apply one or several attribute selection strategies with the appropriate parameters to obtain the number of attributes considered desirable for the new data set.
3) Compare the performance of different strategies, validating by training with the unused sample for the attribute selection strategies and comparing with the complete datasets.
4) Apply expert knowledge or judgment to choose one of the attribute selection strategies. Alternatively, select the subset of attributes resulting from the technique of recursive elimination of attributes with cross-validation, which has shown greater efficacy and validity (Smith, 2018).
5) Develop a test measurement instrument, sticking to the design of the original instrument since it imposes limitations to be respected.
6) From this moment, consider the following steps in the development of the measurement instrument (Mejía Mejía, 2005; Vogt et al., 2012) for the exclusion of attributes or inclusion

of new attributes, or operationalization of new concepts that are necessary and are not included in the original instrument. (Marin et al., 2021)

As we can see in **Table 37**, both ENUT preprocessed datasets have high performance in predicting the two variables of interest, so they are suitable for the task of selecting a subset of attributes by applying the automatic attribute selection tools. Unsurprisingly, the preprocessed ENUT-2019 dataset performs better than its 2014 counterpart since it has more data points than the others. Likewise, we can see that the classifier with the best performance is clearly RF, even though the MLR has the shortest computation time.

Table 37: *Performance metrics, using 10-fold stratified cross-validation, in classification for attributes of the presence of satisfaction in the use of time for academic and work activities of the ENUT-2014 and ENUT-2019 preprocessed datasets, using four different classifiers* (Marin et al., 2021)

Sample	Attribute	Algo.	*Accuracy*	*Precision*	*F_1 Sc.*	C. Time
2014	Academic activities	RF	0.9495	0.9521	0.9494	18.05s
		SVM	0.4977	0.4973	0.4874	46.57s
		DNN	0.8634	0.8758	0.8612	61.45s
		MLR	0.8952	0.8981	0.8950	18.78s
2014	Work activities	RF	0.8149	0.8639	0.8085	0.66s
		SVM	0.5032	0.5037	0.4913	367.47s
		DNN	0.5654	0.6172	0.4745	248.23s
		MLR	0.73772	0.73856	0.73746	1.63s
2019	Academic activities	RF	0.9583	0.9607	0.9583	32.66s
		SVM	0.9128	0.9136	0.9128	37.65s
		DNN	0.9376	0.9378	0.9375	171.18s
		MLR	0.9177	0.9182	0.9177	0.94s
2019	Work activities	RF	0.8142	0.8630	0.8077	130.77s
		SVM	0.7783	0.7896	0.7761	1006.16s
		DNN	0.7507	0.7514	0.7505	1816.61s
		MLR	0.7368	0.7380	0.7365	3.51s

Once the predictive power of the two different datasets has been ascertained, different attribute selection techniques were evaluated, including some already mentioned as not recommended, such as filtering by the highest Pearson correlation index, filtering by univariate f-tests, and forward and backward stepwise regression. Two recommended by the consulted literature were used: recursive attribute elimination and recursive attribute elimination with cross-validation of which the functions from the scikit-learn library were used (RFE and RFECV, respectively). In the case of the last two techniques, which are of the wrapper type, the RF classifier is incorporated since it was the one that had the best performance according to the results in **Table 37**. The results of comparing the attribute selection strategies can be seen in **Table 38**. If the strategy asked for it, a minimum of 30 features were asked as results.

Table 38: *Comparison of performance for classification by the presence of time-use satisfaction in academic and work activities using training subsets obtained by different attribute selection strategies, different classifiers, and training with the same sample used for attribute selection (ENUT-2014). and with a new sample (ENUT-2019)* (Marin et al., 2021)

Selection method	Strategy	Academic Act.		Work Act.	
		Accuracy (2014)	*Accuracy* (2019)	*Accuracy* (2014)	*Accuracy* (2019)
Filter by correlation	RF	0.9576	0.9628	0.8079	0.8070
	SVM	0.9437	0.9516	0.8099	0.8039
	DNN	0.9427	0.9276	0.8082	0.8128
	MLR	0.9154	0.9247	0.7790	0.7818
Filtered, univariate (f-test)	RF	0.9561	0.9641	0.8102	0.8056
	SVM	0.9382	0.9518	0.8083	0.8051
	DNN	0.9432	0.9527	0.8113	0.8115
	MLR	0.9191	0.9216	0.7799	0.7767
Stepping forward	RF	0.9601	0.9642	0.8191	0.8155
	SVM	0.8997	0.9292	0.8034	0.7999
	DNN	0.9439	0.9582	0.8111	0.8144
	MLR	0.9154	0.9246	0.7499	0.7425
Stepping back	RF	0.9583	0.9667	0.8159	0.8156
	SVM	0.8925	0.9412	0.8000	0.8019
	DNN	0.9437	0.9603	0.8051	0.8086

	MLR	0.9202	0.9266	0.7355	0.7346
Recursive elimination	RF	0.9575	0.9619	0.8088	0.8074
	SVM	0.9399	0.9482	0.7927	0.7953
	DNN	0.8922	0.9532	0.7726	0.7797
	MLR	0.9110	0.9203	0.7348	0.7335
Recursive elimination with Cross-Validation	RF	0.9573	0.9639	0.8135	0.8119
	SVM	0.9390	0.9472	0.7936	0.7937
	DNN	0.9480	0.9514	0.7696	0.7812
	MLR	0.9156	0.9202	0.7339	0.7372

In **Table 39**, we can see the time of computation taken by each attribute selection strategy.

Table 39: *Computation times of attribute selection strategies* (Marin et al., 2021)

Attribute selection strategy	Academic act.	Work act.
Filter by correlation	00:00:01	00:00:01
Filtered by univariate tests (f-test)	00:00:01	00:00:01
Stepping forward	00:21:10	00:37:03
Stepping back	00:05:42	00:05:39
Recursive deletion	05:40:28*	03:51:27*
Recursive deletion, cross-validation	21:40:10*	23:05:21*

Format: hh: mm:ss *Using parallel computation.

As seen in **Table 39**, the computation time of the filtering techniques is practically instantaneous, while the stepwise regression takes a few minutes; however, recursive elimination takes several hours and almost a day for recursive delete with cross-validation (with k = 5). The ENUT-2014 dataset with a few hundred dimensions and less than 50,000 records is relatively small compared to typical *Big Data* datasets, so we can see why filtering techniques and stepwise regression are much more used than techniques with a higher level of validity, but that takes much more time to compute in the context of *Big Data*. In the case of the AI-aided design of the measurement instrument for this work, this process is performed only once, and using attribute selection strategies with several hours or even days of computation time can be justified.

As we can see in **Table 37** and **Table 38**, the differences in performance regarding the variables of time-use satisfaction in academic and work activities are minor, both training with the complete datasets and with the subsets suggested by the attribute selection strategies. Likewise, the differences are also small using the suggested subsets, both training with data with the same sample with which the attributes were selected and training them with a different sample, validating the procedure. Not only that but in all cases except one, the performance using the validation sample (ENUT-2019) is not less than that obtained using the original sample (ENUT-2014) with any of the four tested classifiers. This result is attributable to the largest number of records in the new sample and the fact that these are comparable to each other with the same methodological design. Furthermore, as we will see later, the exception is the only case in which a potentially improvable non-automated criterion was applied. However, higher performance is only one of the relevant criteria. The best performance for predicting satisfaction in using time in academic and work activities is seen using stepwise forward regression with random forest, although the approach for attribute selection most recommended by theory is the recursive elimination of attributes with cross-validation, whose difference in performance with stepwise regression is minimal and therefore acceptable.

Regarding the exception mentioned above of lower performance using the ENUT-2019 validation sample compared to using the same ENUT-2014 sample, it occurs when using RF to predict time-use satisfaction in work activities with the set of attributes selected by recursive elimination with cross-validation. Unlike the others, it is essential to note that this method suggests 30 attributes for predicting satisfaction in the use of time in academic activities, which is the number of attributes sought. However, the same strategy suggests 181 attributes for predicting time-use satisfaction in work activities, which indicates that the predictive power for this variable is more distributed among the candidate explanatory variables. Both to be able to continue comparing directly and for practical reasons, and paying a small performance cost for it, the 30 attributes were selected to predict satisfaction in the use of time in work activities in the following way: among the 181 suggested in the case From the recursive elimination with cross-validation, the 27 attributes that said strategy suggested both for the prediction of satisfaction in the use of time in academic and work activities were selected, and three more attributes were selected based on practical considerations from among the remaining attributes suggested for the prediction of satisfaction in the use of time in work activities and that were not suggested for the prediction of satisfaction in the use of time in academic activities.

VI.5.3. Results of the model selection process

Table 40 shows the list of attributes that the process described in the previous section suggested that the new instrument includes for predicting time-use satisfaction in academic and work activities, that is, the data that would be needed from a respondent to predict if they are satisfied with their use of time in academic and work activities.

Table 40: *A subset of attributes of the ENUT (INEGI, 2015, 2020) with the highest performance, produced by recursive elimination with cross-validation, for predicting satisfaction in the use of time in academic and work activities.*

Total number of rooms in the dwelling	Time dedicated to cleaning or tidying up the interior of the dwelling (FS)
Number of people living in the dwelling	Time spent folding/arranging/putting away clothes (ES)
Age in years	Marital status: single/not single
Attends school (age 5 to 24 years)	Time dedicated to sports or physical exercise (ES)
Time spent working (ES)	Time spent watching television without doing any other activity (ES)
Time spent sleeping (includes nap) (ES)	Time spent checking mail, consulting social networks, or (ES)
Time spent sleeping (includes nap) (FS)	Time spent checking mail, consulting social networks, or chatting (FS)
Time spent eating (ES)	The house has a water pump.**
Time spent eating (FS)	Dedicated exclusively to studying
Time dedicated to grooming and personal hygiene; go to the bathroom (en)	Time dedicated to protecting assets and housing (ES)
Time dedicated to grooming and personal hygiene; go to the bathroom (F)	Time spent taking courses or classes (FS)*
Time spent taking courses or classes (ES)	Time spent on tasks, school practices, or study (FS)*
Time spent on tasks, school practices, or study (ES)	Time spent traveling to and from school (FS)*
Time spent traveling to and from school (ES)	Time dedicated to organizing or distributing household chores (ES)**

Time spent cooking or preparing food or drinks (ES)	Time spent consulting information, surfing the Internet (ES)
Time spent serving food; collect; wash; dry and accommodate dishes (EN)	Time spent attending or participating in religious activities/celebrations (ES)**
Time dedicated to cleaning or tidying up the interior of the dwelling (ES)	

**Suggested only for prediction in the context of academic activities.*
***Suggested for prediction only in the context of work activities.*
EN = During the week. FS= During the weekend.

Designing a measurement instrument continues beyond this point as other new attributes can be added.

VI.5.4. Consideration of the methods used

Based on the results in **Table 37** and **Table 38**, it can be seen that the models of 33 attributes selected by the automated attribute selection strategies are validated when training them with a different sample than the ENUT-2019, with which the *accuracy* metrics are not only high but in this case, using the best classifier they are never lower than the performance obtained when training with the complete preprocessed datasets for the ENUT-2019 and ENUT-2014 respectively. This result is evidence that it is possible to use the ENUT surveys to design one with fewer questions that result in microdata with similar or better predictive power for one or several variables of interest, with the advantages that this entails compared to starting the design from scratch.

However, due to the richness of the material presented so far, particularly in Chapter II and this chapter, this approach does not exempt the researcher or analyst from thinking about and studying the concepts operationalized by the original instrument. While this automation procedure necessarily transfers the benefits, limits, and possible errors of the original instrument to the new one, the fact that the ENUT instruments are of superb design and quality minimizes this risk.

VI.6. Interpretable machine learning approach to an intelligent time management tool

In the predictive power assessments of the previous section, the RF classifier has had the best performance relative to the other three tested. While ML models based on decision trees, like RF, are one of the most interpretable ML methods widely available, they have yet to be known for their high accuracy compared with other commonly used ML techniques (Caruana & Niculescu-Mizil, 2006). However, their use has grown in recent years, in part because there are efforts underway to develop ways of constructing decision trees-based models as interpretable surrogate models to explain deep learning ones (Blanco-Justicia & Domingo-Ferrer, 2019), but also because, as we have seen in this chapter, they happen to be a high accuracy option for certain kinds of datasets, such as those produced by measurement instruments designed to explore concepts attributable to people's decisions and adapt well to model the non-linear and complex interactions usually found in these datasets (Dobra, 2009b; Joshi, 2020) even when they contain categorical attributes (Origel-Rivas et al., 2020). These advantages make them a handy option to train and use ML models in social sciences and in industry applications built on datasets from surveys with good statistical design in which these models can have both high accuracy and high interpretability.

The tool developed for this dissertation is a hybrid system that combines interpretable ML and a metaheuristic search algorithm known as genetic algorithms (GA) to inquire into what changes people who do not evaluate themselves as satisfied with their time-use in academic activities can make to their weekly routines to achieve such a state of satisfaction. Besides their interpretability, single decision tree models and RF were chosen for this application because they have been found to be highly accurate for analyzing the datasets used (Marin & Ponce, 2020). Combining them with GA offers ways of presenting suggestions to a user or decision-maker to potentially change how time is used to achieve satisfaction in time-use in academic activities.

Model interpretability, while helpful to describe how the realities represented by the training dataset behave, has another valuable use: the model's parameters can be used to model further how minimally different data points would obtain more desirable results in classification or regression. This use case has essential applications in fields where data points represent realities subject to modification. A general model for an optimization problem involving a single decision tree would look like this:

Parameters:

$a_1, \dots, a_n$ Attributes of a n-dimensional instance (datapoint)

$b_1, \dots, b_m$ Parameters of the trained decision tree with m decision nodes

$k_1, \dots, k_n$ Cost of change for each of the n attributes

Variables:

$x_1, \dots, x_n$ Change in each of the n attributes

Objective function:

$$\min_{x_1,\dots,x_n} f(x_1, \dots, x_n) = \sum_{i=1}^{n} k_i |x_i| \qquad (40)$$

Constraints:

$$[(x_1 + a_1 > b_1) \cap (x_2 + a_2 > b_2) \cap \dots (x_s + a_s > b_s)] \cup$$
$$[(x_1 + a_1 < b_1) \cap (x_3 + a_3 > b_3) \cap \dots (x_t + a_t > b_t)] \cup \dots \qquad (41)$$

As can be seen, the constraints would be the union of all the intersections of the inequations with operators that lead to forming paths to a desirable outcome in leaf nodes $t + 1$, $s + 1$, and so on; therefore, the operators used are not necessarily the original operators of the trained decision tree. The problem can also be seen as several traditional optimization problems, each with a set of straightforward constraints representing a single path down the tree towards the desired result in a leaf node, for example, node $s + 1$:

$$\begin{cases} x_1 + a_1 > b_1 \\ x_2 + a_2 > b_2 \\ \vdots \\ x_s + a_s > b_s \end{cases} \qquad (42)$$

Also, the problem can be taken as a global one in which all the paths are considered at once, and the whole trained decision tree $\mathbb{T}$ can be used as a constraint in which d delimiters the set of desired outcomes or is the desired outcome:

$$\mathbb{T}(x_1, \dots, x_n) > d \quad \text{or} \quad \mathbb{T}(x_1, \dots, x_n) < d \quad \text{or} \quad \mathbb{T}(x_1, \dots, x_n) = d \qquad (43)$$

Furthermore, the constraints could be relaxed by trimming parts of the intersections at the level that the impurity or error of the nodes is considered tolerable to the decision-makers or by eliminating specific paths down the decision tree to forestalling them. These judgments can be made iteratively, or an optimization hybrid metamodel can be designed to consider multiple options of constraint relaxation and path elimination.

A simple example of this approach can be obtained using a trained classification decision tree with the well-known Iris dataset (Dua & Graff, 2017), as seen in **Figure 39**.

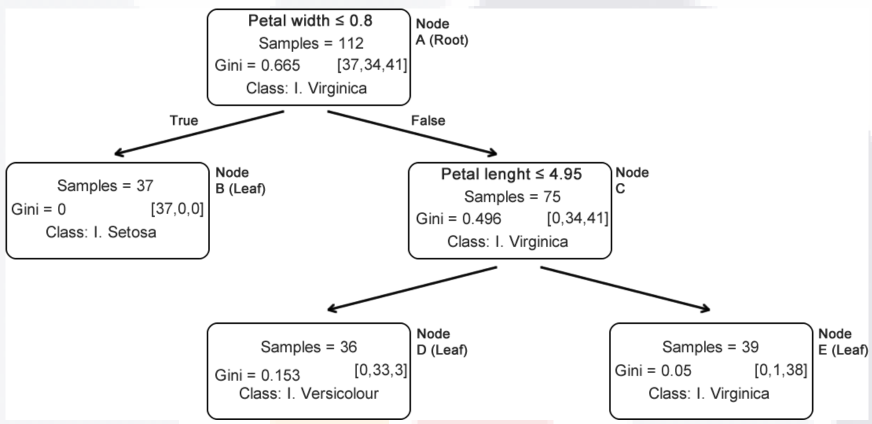


Figure 39: *A decision tree trained with attributes belonging to the Iris dataset*

If, by some process of domestication, the plants were modified in petal width and lengths so that they all look like the original Iris Virginica, the objective function in which x_1 stands for petal width and x_2 for petal length, would look like this:

$$\min_{x_1,x_2} f(x_1, x_2) = k_1|x_1| + k_2|x_2| \tag{44}$$

Apart from feasibility constraints, the main constraints of the model would use the inequality operators that lead to the desired outcome as in:

$$(x_1 + a_1 > 0.8) \cap (x_2 + a_2 > 4.95) \tag{45}$$

However, if, according to a decision-maker, the domesticated plants must be like the original Iris Virginica only in petal width, then the primary constraint would consider the branch with the Iris

Virginia leaf node only down to Node C in the diagram in Fig. 1, even as that node has a high impurity as far as the original decision tree model goes. The relaxed constraint would look thus:

$$(x_1 + a_1 > 0.8) \tag{46}$$

VI.7. Genetic algorithm subjective wellbeing enhancement

The results in **Table 37** suggest that a decision tree trained with either of the ENUT surveys' microdata will have enough accuracy to both explain to students how their time use patterns lead to satisfaction or dissatisfaction in time use in academic activities and to power an optimization model to offer suggestions to students that the model finds currently dissatisfied with their time use in such settings. However, a practical problem arises. The original ENUT surveys contain hundreds of questions, and their use as-is is impractical. However, the briefer attribute set in TABLE can be used instead.

A specific model for an optimization problem testing the decision tree, trained only the time-use attributes in **Table 40** and eight more time-use attributes deemed as relevant to include by the research team, to get users from unsatisfied to satisfied with time use in academic activities with minimum change to weekly routines is:

Parameters:

$a_1, \dots, a_{33}$	Time use attributes of the instance (users or respondents)
$b_1, \dots, b_{2181}$	Parameters of the decision tree's 2181 decision nodes
$k_i = 1, \forall\, i \in [1,24]$	Cost of change for each of the 24 attributes is unitary

Variables:

$x_1, \dots, x_{33} \in [-60,60]$	Change in each of the n time use attributes (minutes)

Objective function:

$$\min_{x_1,\dots,x_{24}} f(x_1, \dots, x_{33}) = \sum_{i=1}^{33} |x_i| \tag{47}$$

Constraints:

$$[(x_1 + a_1 > b_1) \cap (x_2 + a_2 > b_2) \cap \dots (x_s + a_s > b_s)] \cup$$
$$[(x_1 + a_1 > b_1) \cap (x_3 + a_3 > b_3) \cap \dots (x_t + a_t > b_t)] \cup \dots \tag{48}$$

Or:

$$\mathbb{T}(x_1, \dots, x_{33}) > 0 \tag{49}$$

Where $\mathbb{T}$ is the complete trained decision tree, this constraint is equivalent to the whole set of intersections for the unrelaxed constraints created with the parameters of the decision tree. An output value of 1 for the decision tree means the presence of satisfaction in time-use in academic activities, while a value of 0 means its absence.

As can be inferred from the above model, instances with 33 attributes can have a vast number of combinations of changes, considering that time attributes can be changed to go up or down by different amounts. Also, with 2181 nodes, hundreds of intersections represent possible ways from the root of the decision tree to a leaf node that concludes with the instance being classified as with satisfaction with time use in academic activities. This complexity calls for a metaheuristic method, with evolutionary approaches suitable. In this case, a single run of a genetic algorithm was used, but multiple runs of these and the decision tree algorithms can be used for a more robust analysis.

For this problem, a genetic algorithm (Sammut, 2017) was used with a population of 200, with probabilities of 0.90 for recombination, 0.05 for mutation, and 200 generations. The selection was random and by contest. Recombination was done by erasing the values of an instance's vector from a randomly selected point and appending to the severed vector the missing part from another randomly selected instance. The mutation was done by assigning a value of 0 to a randomly selected variable in the instance selected for mutation. All random processes were done using uniformly distributed random variables. Factibility checks were performed after every operator by checking that the sum of all variables does not exceed 1200 minutes or is equal to zero, and discarding those that meet these conditions and substituting them for new individuals randomly created by the same process by which the initial population was created with values randomly chosen from a range of minus 50 to 50 weekly minutes, just as other random values are assigned. A diagram of the cycle for the algorithm is shown in **Figure 40**.

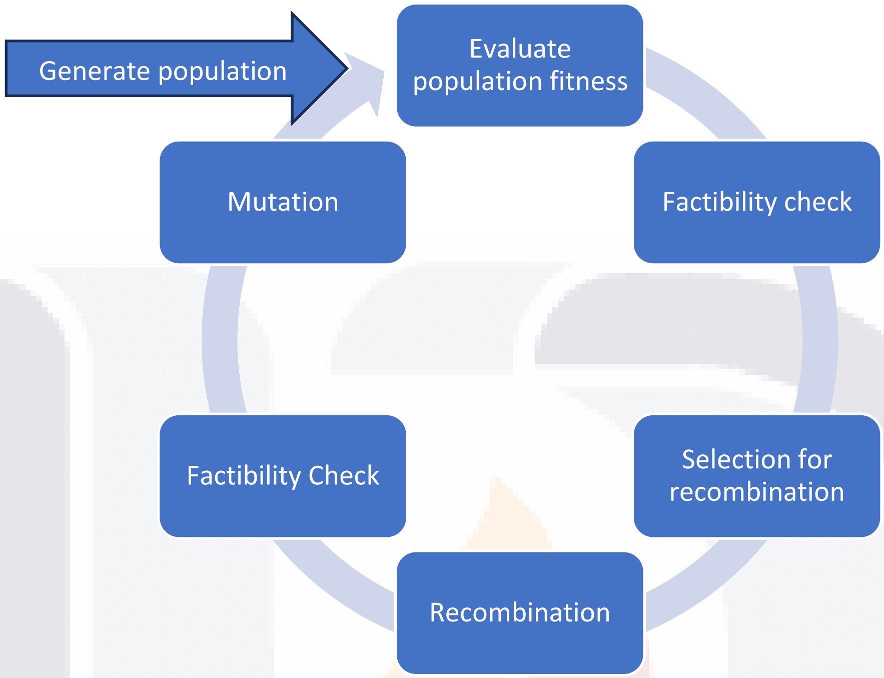


***Figure 40**: Cycle of the genetic algorithm*

VI.8. Survey and feedback design

The intelligent time management tool tested in the experiment described in the following chapters has at its heart a genetic search algorithm looking for the minimal changes to bring a respondent from time-use dissatisfaction to satisfaction in academic activities. However, the front face of this tool is an electronic survey in the form of a web app that can contain the questions corresponding to the attributes in **Table 40** and others that can bring information about wellbeing in terms of agency. The web app was developed using Hypertext Markup Language 5 (HTML5), Cascading Style Sheets 3 (CSS3), and PHP Hypertext Preprocessor, and it redirects the respondent's data over a Secure socket layer (SSL) connection to a MariaDB encrypted database, from which the research team can download the data as a comma-separated values (CSV) file.

This web app has been used to produce one electronic survey to collect the data relevant to the research and to send the tool's feedback or recommendations to the users via e-mail.

VII. Experimental design and implementation

While most of the material reviewed in this dissertation deals with empirical data drawn from the microdata of national surveys from nationwide representative samples, the benefits of the approach on human wellbeing and agency are separate from these data. Therefore, producing a research dataset through an intelligent time management tool is warranted. While we have presented essential results from the analysis and processing of the ENUT microdata, we have yet to show if these results, although already validated from two different samples (2014 and 2019 ENUT microdata), can be generalizable enough to be replicated in a particular less-than-representative sample taken in the kind of practical context in which intelligent time management tools would be applied.

The practical context in which the experimental contributions are deployed is essential. It would hardly be useful to design an experiment with a sophisticated method to construct a highly representative and random sample if we have already validated the results we want to replicate with two such samples and, more importantly, the actual use of the intelligent time management tools is with groups of high homogeneity, such as the members of workshops, corporate departments, groups of students, and other kinds of purposeful assemblages of people in which there is a need to better time management, performance and human wellbeing in a structured way.

In other words, it is not the case that, at this point, we are still determining the evidence that the intelligent time management tools would have accurate results while working with a representative sample from the general population. We are now interested in these tools working accurately, or even more accurately, with a sample representing a set of groups we are interested in thoroughly in their rather specific context. For this reason, we decided to use convenience samples and simple analysis methods to let the results speak more plainly (Montgomery, 2020). Thus, while done to scientific and statistical standards, the experiment structure presented in this chapter is done with practicality and applicability in mind, as well as with the intention of generation a dataset of student data amenable to further research.

The experiment was carried out with several groups of students in the Basic Sciences Centre of the Aguascalientes State Autonomous University (UAA). The decision to carry out the experiment with the help of the students of the Basic Sciences Centre obeyed two main reasons: the accuracy of the classification algorithms is better with students than with employees, as was shown in **Table 23**, and the costs and timetables were less onerous in an academic context than in a business organization context. Nonetheless, as has been remarked, students are among the groups that employ time management the most, given their structured class days and calendars.

VII.1. Hypothesis and research questions in practice

As seen in Chapter I, the main hypothesis is that intelligent time management tools can accurately predict time-use satisfaction significantly better than a random classifier with a uniform distribution and a test using only the training dataset under a 10-fold cross-validation regime. This hypothesis is born from our main research question:

> In college students, can an intelligent time management tool trained with mostly time-use datasets and tested with an experimental dataset predict significantly better the level of time-use satisfaction in academic activities than a random classifier and a cross-validation regimen using only the training dataset?

Other complementary research questions can address all other variables of time-use satisfaction contained in the ENUT microdata; however, the SWB component relevant to this experiment is linked to students' use of time in their academic activities, and only then, through the properties of time-use satisfaction, to other variables of SWB (Boniwell, 2006). All other changes in SWB are considered incidental to the changes recommended by the intelligent time management tool focused on improving time-use satisfaction in academic activities for a group of college students. Hence, the main research questions deal with time-use satisfaction in academic activities.

VII.2. Experimental design

To test the main hypothesis and answer the main research question and all complementary research questions, an intelligent time management tool called *Wellbeing and Agency Intelligent Time Management System 2022 Version* (WAIT-2022) was designed with a front-facing user interface in the form of a web app that presents the users with an initial survey.

After taking the initial survey, WAIT-2022 processes users’ answers and sends them an e-mail after a number of days with system feedback in the form of suggestions for time use change: at most five suggestions to decrease time allocated to some activities, and five suggestions to increase allocation to other activities along with some complementary advice. Not all users get personalized recommendations from WAIT-2022, though. WAIT-2022 can only process personalized recommendations for users it predicts do not currently have time-use satisfaction in academic activities, even if they say they do have it. However, paying respect to the subjectivity of assigning values to the research variables, WAIT-2022 only processes and sends personalized recommendations to users that state they do not have time-use satisfaction in academic activities and are predicted not to have it by the system. All users who do not get personalized recommendations get generic recommendations based on the following:

1. The most frequent recommendations presented by WAIT-2022 to a random group of 400 ENUT-2014 respondents who are students.
2. The most frequent recommendations presented by WAIT-2022 to their peers that did get personalized recommendations.
3. Generic recommendations from Gallup's Wellbeing Five model (Rath & Harter, 2010), as shown in **Table 1**.

On the other hand, users who get personalized recommendations get a summary of, at most, the five increases and five decreases of time allocations to activities that get them the nearest to time-use satisfaction in their academic activities. They also get the generic recommendations based on Gallup's Wellbeing Five model.

VII.2.1. Survey

The front-facing user interface of WAIT-2022 is in the form of a web app that presents the users with a survey that contains all the questions linked to the attributes in **Table 40** that form part of the abbreviated ENUT-based measurement instrument presented in the previous chapter. Given the review of the wellbeing literature, particularly the emphasis of the Wellbeing Five model on social and community activities not included in the attributes in **Table 40** five more activities were included in the WAIT-2022 initial survey to enlarge the training subdataset with the attributed linked to these activities in the ENUT microdata. The additional activities are:

1. Carry out artistic or cultural activities (playing a musical instrument, painting, performing plastic, graphic, literary, or performing arts, including dance).
2. Attend civic or political events (meetings, rallies, parades, marches, celebrations, etc.).
3. Volunteer or participate in organizations such as civil associations, the Red Cross, nursing homes, homes, DIF, hospitals, churches, Alcoholics Anonymous, political parties, etc.
4. Attend or participate in religious activities or celebrations.

5. Dedicate special time (without doing other activities) to the members of your household to talk about daily activities, console, or advise.

The addition of the previous five questions meant ten additional attributes, given that time-use allocations are required for both weekdays and weekends. The text of all the questions in the WAIT-2022 initial survey is lifted directly from the original ENUT surveys, with both surveys in Spanish.

Also, binary yes/no questions related to agency-based wellbeing were added to each question about time use in a particular activity, related to the factor seen in Chapter II: It improves my capacity or ability to interact with my environment and change it if necessary.

The procedure to create a briefer measurement instrument presented by the research team in a earlier publication (Marin et al., 2021) and presented in the previous chapter was followe; this means that the good experimental design of the original ENUT survey is inherited by the initial survey of the WAIT-2022 insofar as it pertains to the text of the questions and the attributes linked to it, and therefore to the relevance and pertinence of these variables to be processed by classifiers trained with ENUT microdata. However, some of the original design is inevitably lost due to the electronic presentation and, of course, the different order of the questions and the added questions via expert knowledge. This situation is the case with almost all brief measurement instruments this methodology can produce, and this is why it is vital to test that predictive power is enhanced or at least conserved with the new sample.

The questionaries for the survey, as originally written in Spanish and a translation into English, can be found in Annex 1 and Annex 2, respectively. Similarly, the text of the e-mail sent to students as originally written in Spanish and a translation into English can be found in Annex 3 and Annex 4, respectively.

VII.2.2. Model implementations

WAIT-2022 uses implementations of the following intelligent tools presented in the previous chapter.

- Brief measurement instrument with AI-aided design
- Interpretable classifiers based on decision trees
    - Single decision tree algorithm
    - Random forest algorithm

- GA for SWB enhancement

These pieces of the WAIT-2022 tool are used as already presented or in its standard form in the case of the genetic search algorithm (Sammut, 2017; Sivanandam S.N. & Deepa, 2008), as follows.

The brief measurement instrument with AI-aided design obtained from the list of attributes in **Table 40** plus the five added activities' questions via expert knowledge and agency questions is implemented as a survey in the WAIT-2022 web app, which is HTML5/CSS3 standards compliant and stores the raw data obtained from its surveys in a secure MariaDB database via an encrypted secure socket layer (SSL) connection. The users can use the WAIT-2022 web app via desktop or mobile web browsers such as Firefox, Chrome, Edge, Opera, Vivaldi, Safari, or any other standard-compliant modern web browser. Then, this data is preprocessed to be fed to the interpretable classifiers based on DTs, which then produce a single trained tree that can be evaluated to classify into different classes of SWB variables, but in this particular case, time-use satisfaction with academic activities.

First, an RF model is trained with the ENUT-2019 preprocessed dataset, and the single DT with the greatest accuracy is selected and compared with a single DT trained with the same ENUT-2019 preprocessed dataset. Whichever of these has the greater accuracy has its parameters used to feed the genetic search algorithm. This genetic search algorithm is used to search for the minimum changes in time use allocation that improve the SWB class of the user, in the case of WAIT-2022 users, the time-use satisfaction in academic activities. Then, these changes are separated into sets of changes that indicate decreases and increases in time-use allocations. They are then ordered by how near to being classified as having time-use satisfaction in academic activities get them in the trained decision tree used for the search parameters. From each set, the subset of, at most, the first five changes is sent to the user as a change recommendations summary along with additional generic recommendations.

A diagram of the process explained in the above paragraph is shown in **Figure 41**.

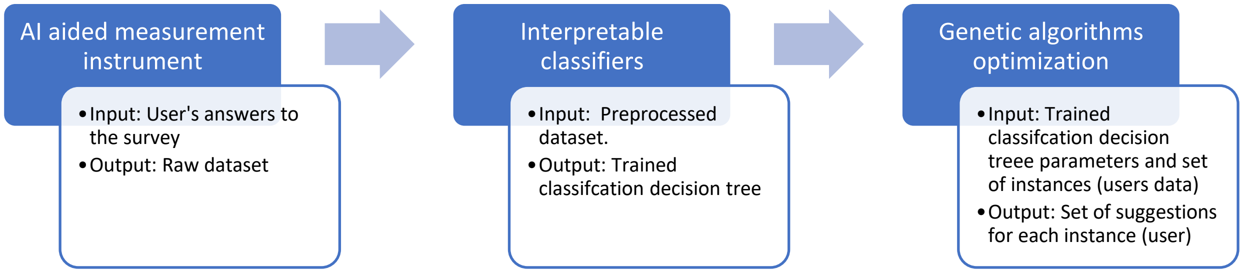


Figure 41: *Diagram of data flow on the constituent intelligent algorithms of WAIT-2022 tool*

The AI-aided measurement instrument and the interpretable classifiers used were as described in the previous chapter, while the genetic algorithm used was a standard implementation with simple operators (Sivanandam S.N. & Deepa, 2008) with the parameters shown in **Table 41** and already discussed in the previous chapter.

Table 41: *Parameters used for the standard genetic search algorithm configuration*

Parameter	Value
Population	200
Inferior limit for values	-50
Superior limit for values	-50
Probability of crossover	0.90
Probability of mutation	0.05
Maximum number of iterations	200

The narrow inferior and superior limits for the recommendations shown in **Table 41** should be noted. This narrow search space is adequate given that the parameters of a DT trained with mostly time-use data are usually tight, and it is not necessary to make significant changes to shift the decision at any particular node. However, this means that most suggestions to increase or decrease time allocations to particular activities are not accompanied by numeric suggestions in minutes but are instead of the “increase/decrease time allocated in a reasonable quantity” type. A feasibility check was performed after every application of an operator (recombination and mutation) to check that the sum of all changes suggested to a particular user does not exceed 1200 minutes or is equal to zero.

VI.2.3. Experiment sample, methodology, and calendar

The experiment was done with the voluntary participation of undergraduate students from the Basic Sciences Centre of the Aguascalientes State Autonomous University (Universidad Autónoma de Aguascalientes) in Aguascalientes, México. The sample was a convenience sample of six groups from different majors: mathematics and intelligent computation. The sample was 295 students, with only 227 students accepting to participate and answering the survey to completion. The surveys were applied from November 1 to December 9, 2022. Each group was asked to participate in the experiment during a preprogrammed class in which desktop computers were available to all students, allowing those who did not agree to participate to decline to do so discretely and do other activities on the computers. Wireless internet access was available, and a small number of students preferred to answer the survey on their cellphones as the WAIT-2022 survey web app was compatible with mobile browsers. At least two research team members were physically available during this phase of the experiment.

The experiment session was 50 minutes long and consisted of four parts: a presentation, a brief introduction to SWB and agency themes, time to read the survey, ask questions, and answer the survey, and a time for final remarks. A detailed description of these phases is shown in **Table 42**.

***Table 42**: Description of the phases of the WAIT-2022 survey session*

Phase	Duration	Done by	Description
Presentation	5:00 min	Class professor	The class professor would introduce the member(s) of the research team, tell the group about the project, and ask them to participate, promising feedback from the AI tool shortly.
A brief introduction to SWB and agency	11:00 min	Member of the research team	A brief presentation on the themes of wellbeing and agency would be presented to the group, as well as remarks about their data privacy if they choose to participate.
Time to read the survey's instructions	32:00 min	Students	The survey's instructions would be read out loud in their entirety so that

and answer the survey			questions could arise. Then, those who chose to answer the survey would do so. The research team members answered individual questions during this time.
Final remarks	2:00 min	Member of the research team	The research team members shared their e-mail, indicated to students that they would receive feedback from WAIT-2022, and said goodbye to the group.

The research team and its objectives were introduced during the presentation, and students were assured of complete privacy regarding their data. All technical questions were answered promptly and with the assistance of a research team member. All questions about clarity were answered with a variation of "Please, interpret the question as you see fit or as it makes sense in your life." The survey's instructions were read first, and it was explained that all the answers were personal and subjective, that is, that the only relevant opinion and criterium were the respondent's.

Finally, during the final remarks phase, the group would be reminded that they would receive feedback from the WAIT-2022 tool in the coming weeks. The recommendations from WAIT-2023 were sent on January 8, 2023.

VII.3. Tests

As already mentioned, the applied nature of the research presented in these experimental contributions makes it important to use the most straightforward methods available that are appropriate given the nature of the data and those methods already widely used in the field of ML. For this reason, t-tests, correlation analysis, and cross-validation are used in most tests.

VII.3.1. Variables of interest

The variables of interest are those that are going to be used for statistical tests and result's compilation, as well as for testing the hypotheses presented in Chapter I. These variables and their characteristics are shown in **Table 43**.

__Table 43__: Variables of interest, their possible values, and their relevance to the tests performed

Variable	Values	Relevant in testing
Time-use satisfaction in…		
Academic activities	{0,1}	Main hypothesis.
Social/Family activities	{0,1}	Supplementary hypotheses.
Work	{0,1}	
Commutes	{0,1}	
Support close ones	{0,1}	
Pay/Paperwork	{0,1}	
Preferred activities	{0,1}	
Agency-based wellbeing while doing…		
Sleep	{0,1}	Proposed wellbeing definition.
Personal hygiene/care	{0,1}	
Eating meals	{0,1}	
Classes	{0,1}	
Social events	{0,1}	
Sports and workouts	{0,1}	
Conversation with close ones	{0,1}	
Homework	{0,1}	
Cleaning (house's interior)	{0,1}	
Reading (not for studies)	{0,1}	
Art and cultural activities	{0,1}	
Social networks, e-mail, chat	{0,1}	
Watch movies, streaming, soap-operas	{0,1}	
Cook and prepare meals and beverages	{0,1}	
Separating, folding, and storing clothes	{0,1}	
Serve food, cleaning, washing dishes	{0,1}	

Commutes (to university)	{0,1}
Work	{0,1}
Close and secure house and other goods	{0,1}
Civic and political activities	{0,1}
Voluntary activities	{0,1}
Religious activities	{0,1}

VII.3.2. Testing the main hypotheses and validating the WAIT-2022 accuracy

Given that the initial survey of the WAIT-2022 provides labeled data, a predictive power analysis was done using the WAIT-2022 DT-based classifiers trained with the ENUT-2019 preprocessed dataset using as test dataset the one obtained from the WAIT-2022 initial survey.

First, a test is to be done to determine that the average accuracy from using a DT-based classifier trained, specifically an RF model, with ENUT-2019 data under a 10-fold cross-validation regime is significantly greater than the accuracy of a random classifier with a uniform distribution, which in the case of a binary variable of interest is 0.5000. The hypotheses involved in such a test are:

> **H_0:** The average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime is equal to or less than that of a random classifier with a uniform distribution.
>
> **H_1:** The average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime is greater than that of a random classifier with a uniform distribution.

Given that we are also interested in the WAIT-2022 initial survey conserving the accuracy of the original ENUT-2019 cross-validations regimes, we are also interested in testing if the accuracy of using the WAIT-2022 as test dataset in a DT-based classifier trained with ENUT-2019 is greater than that of the average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime. The hypotheses involved in such a test are:

H_0: The accuracy from testing the WAIT-2022 initial survey dataset on a DT-based classifier trained with ENUT-2019 is equal to or less than the average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime.

H_1: The accuracy from testing the WAIT-2022 initial survey dataset on a DT-based classifier trained with ENUT-2019 is greater than the average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime.

If both previous null hypotheses are rejected, we can conclude that the ENUT-2019 cross-validation regime has better accuracy than a random classifier with a uniform distribution and that testing the WAIT-2022 initial survey dataset on a decision tree-based classifier trained with ENUT-2019 data has greater accuracy than the ENUT-2019 cross-validation regime. Therefore, the WAIT-2022 initial survey test would have an accuracy that is better than a random classifier with a uniform distribution and is, in fact, better than a simple test of accuracy involving only its training dataset employing cross-validation. This result would supports our main hypothesis that intelligent time management tools can accurately predict SWB levels while training with mostly time-use datasets better than random classifiers and better than just testing their training datasets under cross-validation regimes.

VII.3.3. Complementary research questions

The complementary research questions, linked to other specific objectives, were posited as follows:

- Can basic principles of preserving and enhancing well-being be articulated to develop AI tools?
- Can time-use survey data be characterized as human behavior in an economy of time?
- Can interpretable machine learning algorithms trained mainly with time-use data be used for prediction and recommendations in wellbeing?

In the following chapters, the research team will argue if and why these research questions were answered.

VII.4. Correlation analysis of agency-related variables

Finally, including an agency-based wellbeing variable in the initial survey allows for a correlation analysis between this variable and the SWB attributes included in the ENUT-2019 and ENUT-2014 preprocessed datasets. The Pearson coefficient was used for the correlation analysis, as recommended for this kind of data in the field literature (Ark, 2014).

Since these attributes are not represented in the training datasets, no other tests are performed using them.

VIII. Results and discussion

The implementation of WAIT-2022 produced one dataset from its initial survey designed to be tested in its DT-based intelligent classification algorithm trained with ENUT-2019 microdata. Also, data was produced about the accuracy of WAIT-2022 predictions of the presence of time-use satisfaction, the links of the agency-based wellbeing variable to other SWB variables, recommendations given by WAIT-2022 to students unsatisfied with how they use their time in their academic activities, and general advice to those that already have time-use satisfaction as students or are predicted to have it.

VIII.1. Results from the WAIT-2022 initial survey

The dataset produced by the initial survey was compared to ascertain the differences between the sample of students using WAIT-2022 and the general population sample that took the ENUT-2019 survey in terms of time-use satisfaction. No data on other SWB variables are compared, given the low performance of the algorithms used to construct WAIT-2022 with these other variables. These results are shown in **Table 44**.

Table 44: *Results of the WAIT-2022 initial survey for time-use satisfaction compared to the ENUT-2019 survey*

Activities in which the time is used	Users with time-use satisfaction (Number and percentage)				Difference
	WAIT-2022 initial survey		ENUT-2019 survey		
Academic	80 users	35.24%	8606 users	12.05%	23.19%
Social/Family	74 users	32.60%	23821 users	33.36%	-0.76%
Work	44 users	19.38%	27834 users	38.98%	-19.60%
Commutes	82 users	36.12%	33877 users	47.44%	-11.32%
Support close ones	92 users	40.53%	20409 users	28.58%	11.95%
Pay/Paperwork	56 users	24.67%	9859 users	13.81%	10.86%
Preferred	40 users	17.62%	30935 users	43.32%	-25.70%
Domestic	106 users	46.70%	41953 users	58.75%	-12.05%

Comparing the results in **Table 44**, the most dramatic difference between the ENUT-2019 and the WAIT-2022 sample is in time-use satisfaction in academic and preferred activities. The first difference will be explained later in this chapter by virtue of most people who are students being satisfied with their time allocation to academic activities, and most people who are not students having no satisfaction with their time allocation (or lack thereof) to academic activities whether because they are unsatisfied that they do not allocate time to it or just do not make such cognitive judgment which also translates as lack of time-use satisfaction in academic activities, just not dissatisfaction. On the other hand, the staggering difference in which the students of the WAIT-2022 sample are 25.70% less satisfied with their time use in their preferred activities cannot be explained directly by data contained in the ENUT-2019 dataset, as the preferred activities are not specified.

Also, we are interested in showing the relationship between the agency-based wellbeing variable included in the WAIT-2022 initial survey and the other SWB attributes included in it. This agency-based wellbeing variable is the one that will serve to test whether the wellbeing definition posited by the research team in Chapter II has a link with the robust SWB construct. The proportions of respondents who agree with the proposition that mirrors our proposed definition of wellbeing about each activity are shown in **Table 45**, and these results show some interesting links.

Table 45: *Results of WAIT-2022 initial survey ordered in terms of the number of people that, for each activity, they find true each of the three statements related to agency*

Activities	Improves my capacity…[1]		Ad-hoc classification
Sleep	135	59.47%	Self-care
Personal hygiene/care	117	51.54%	
Eating meals	116	51.10%	
Classes	112	49.34%	Balanced wellbeing life
Social events	100	44.05%	
Sports and workouts	96	42.29%	
Conversation with close ones	95	41.85%	
Homework	89	39.21%	Everyday activities
Cleaning (house's interior)	88	38.77%	
Reading (not for studies)	86	37.89%	
Art and cultural activities	70	30.84%	
Social networks, e-mail, chat	60	26.43%	
Watch movies, streaming, soap-operas	55	24.23%	

Cook and prepare meals and beverages	54	23.79%	
Separating, folding, and storing clothes	51	22.47%	
Serve food, cleaning, washing dishes	47	20.70%	
Commutes (to university)	43	18.94%	
Work	36	15.86%	
Close and secure house and other goods	33	14.54%	
Civic and political activities	32	14.10%	Social engagement activities
Religious activities	30	13.22%	
Volunteer activities	28	12.33%	

[1] *Improves my capacity or ability to interact with my environments and change them if necessary.*

Some interesting results in **Table 45** are that the most highly rated activities in terms of agency-based wellbeing are those that, in an ad-hoc classification judgment, can be considered self-care: sleeping, meals, and personal hygiene and care, with above 50% of respondents considering these activities allow them to increase their ability to interact with and change their environments. Under 50% but above 40% are activities that could very well be part of a balanced life of wellbeing according to the Wellbeing Five model shown in **Table 1** such as classes (learning), social events, sports and workouts, and conversations with close ones. Between these highly rated activities and the lowest ones are a variety of domestic and entertainment activities relevant to a student's life but, arguably, not essential. Finally, the lower-rated activities are those collectively considered the most unimportant to agency-based wellbeing by the students, such as securing their house, work, civic and political activities, religious activities, and, least of all, volunteer activities.

It is crucial to notice the importance of the role of the environment in these results to ascertain what they do not say. For example, while the Wellbeing Five model (Rath & Harter, 2010) stresses the importance of community participation for a life of wellbeing of which voluntary, civic, and political activities can be an essential part, the low rating of these kinds of activities in terms of how they help students to develop their capabilities to interact and change their environments are insufficient to judge whether this means disinterest to better their social and participatory capabilities on the part of the students or a failure of the kind of activities available to them to actually be that kind of agency-based wellbeing source.

VIII.2. Accuracy validation

Before proceeding to present data from the recommendations given to students by WAIT-2022, as well as statistical tests and validations, the issue of accuracy should be dealt with first. As mentioned in previous chapters, the algorithms' accuracy guarantees their real-life usefulness, as an inaccurate algorithm will provide parameters to the search algorithm of WAIT-2022 that will generate a search space with no actual relationship to the reality of effective recommendations to enhance SWB levels. Therefore, a predictive power analysis of accuracy using a DT-based classifier trained with the ENUT-2019 preprocessed dataset and using as a test dataset the one generated by the WAIT-2022 initial survey should find that the accuracies are higher than those of a similar predictive power analysis but using only the ENUT-2019 preprocessed dataset as a source for both test and training dataset under a 10-fold cross-validation regime. This condition certainly is a high bar, but it can be proposed because the test by cross-validation has already been validated with two different samples (ENUT-2019 and ENUT-2014), and a more compact test dataset composed of relevant attributes from a model selection process also validated with two samples, should perform slightly better.

In the case of the ENUT-2019 10-fold cross-validation regime, only students are included in the analysis. The reason for not using the complete ENUT-2019 preprocessed dataset in this evaluation is easy to show by taking advantage of the visual interpretability of the tree-based algorithms. This interpretability makes it possible to see at least the decision tree's first levels graphically with clarity. As seen in **Figure 42** and **Figure 43**, early in the decision tree process, most people satisfied with their time use in academic activities are correctly classified just by the fact of being students. This circumstance, added to the WAIT-2022 surveys being applied only to students, makes this a fairer comparison.

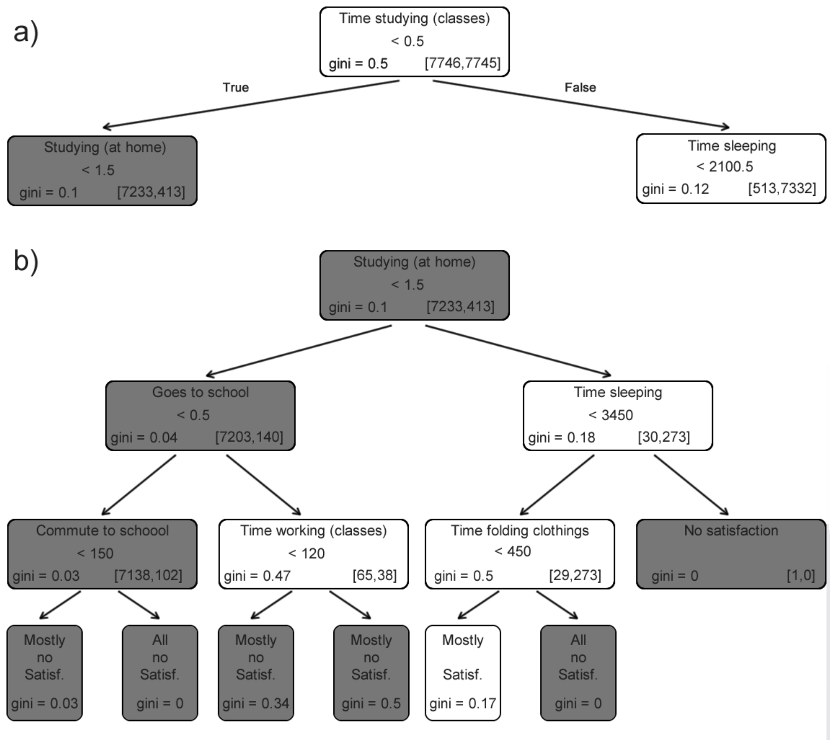


Figure 42: *Diagram of the root and upper part (a) and leftmost part (b) of the decision tree trained with the ENUT-2019 preprocessed dataset used to power WAIT-2022. Gray nodes are dominated by the absence of time-use satisfaction in academic activities, while its presence dominates white nodes.*

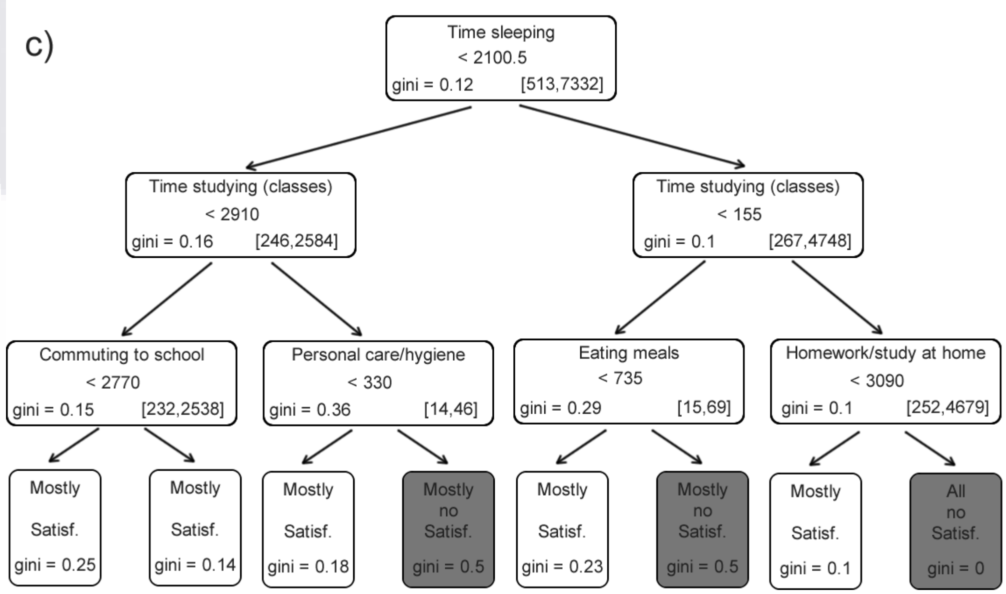


Figure 43: *Diagram of the rightmost part (c) of the decision tree trained with the ENUT-2019 preprocessed dataset used to power WAIT-2022. Gray nodes are dominated by the absence of time-use satisfaction in academic activities, while white ones by its presence*

The accuracy validation for time-use satisfaction in academic activities should guarantee that the results obtained testing the dataset produced by the WAIT-2022 initial survey with a decision tree trained with the ENUT-2019 preprocessed dataset will be greater in accuracy than the 10-fold cross-validation regime using only the ENUT-2019 preprocessed dataset and seen in **Table 37**, which is a performance already validated with a comparison with the ENUT-2014 preprocessed dataset which originates from another sample altogether. The results of the accuracy analysis are shown in **Table 46**.

Table 46: *Predictive analysis of accuracy for time-use satisfaction variables training and testing with the ENUT-2019 and ENUT-2014 complete preprocessed datasets and with only the subset that is students under a 10-fold cross-validation regime using an RF classifier; comparing with testing the WAIT-2022 with an RF classifier trained with the ENUT-2019 subdataset containing only the attributes collected in the WAIT-2022 initial survey that are contained in the ENUT survey*

Type of activities evaluated	ENUT-2014 complete*	ENUT-2019 complete*	ENUT-2019 only students*	WAIT-2022 Initial Survey
Domestic	0.6036	0.6335	0.6005	0.4758
Academic	0.9608	0.9600	0.5678	0.6035
Work	0.8097	0.8167	0.9177	0.7621
Support to close ones	0.6179	0.6945	0.6781	0.5947
Social/family	0.7036	0.6975	0.6707	0.6696
Commutes	0.7209	0.7816	0.6432	0.6872
Payments/paperwork	0.9582	0.9540	0.8997	0.7533
Preferred	0.5479	0.6070	0.5716	0.8194

**Under a 10-fold cross-validation regime*

One interesting result in **Table 46** is the enhanced accuracy for the variable of time-use satisfaction in preferred activities, which soars to 0.8194 from 0.6070 with the ENUT-2019 10-fold cross-validation tests, which indicates that this particular variable is more easily predicted in a homogenous student group or sample than in a set of students from a nationwide representative sample, which may indicate some particular situation with the sample of students that used the WAIT-2022 which hinders their satisfaction on how they allocate their time to their preferred activities.

As far as the variable of interest goes, which is time-use satisfaction in academic activities, we can see in **Table 46** that while the accuracy of the 10-fold cross-validation regime using only the ENUT-

2019 preprocessed dataset is much higher than the accuracy testing the WAIT-2022 initial survey dataset, a fairer comparison is to the student sub-dataset from the ENUT-2019, for the reason explained in the previous paragraphs and shown in **Figure 42** and **Figure 43**. In fact, the performance is slightly better for the WAIT-2022 initial survey dataset, something expected after discarding hundreds of less relevant attributes in the process of model selection validated with two different samples, as was done in this instance with the ENUT-2014 and ENUT-2019 preprocessed datasets and adding five activities to the list of attributes via expert knowledge. This very process of model selection caused the bigger increase in accuracy for time-use satisfaction in preferred activities.

Using a 10-fold cross-validation regime, we can perform a one-tailed Student's t-test using the results from each fold, shown in **Table 47**, to test the main hypotheses.

Table 47: *Results of the accuracy calculation for each of the 10 folds of the cross-validation regime for the classification of time-use satisfaction in academic activities using the ENUT-2019 preprocessed dataset*

0.5808	0.5721	0.5753	0.5939	0.5677
0.5644	0.5415	0.5731	0.5655	0.5437

Using the data in **Table 47**, we can perform a one-tailed Student's t-test with the formula:

$$t = \frac{m - \mu}{s/\sqrt{n}} \quad (50)$$

Where:

t = Student's t

m = mean

μ = theoretical value

s = standard deviation

n = variable set size

The first test, at 99% confidence, with $n - 1$ degrees of freedom, is performed with the following hypotheses in mind:

H$_0$: The average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime is equal to or less than that of a random classifier with a uniform distribution. This hypothesis is expressed as follows:

$$m \leq 0.5000 \tag{51}$$

H$_1$: The average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime is greater than that of a random classifier with a uniform distribution. This hypothesis is expressed as:

$$m > 0.5000 \tag{52}$$

The calculations are as follows:

$$t = \frac{0.5678 - 0.5000}{0.01579 \,/\, \sqrt{10}} = \frac{0.0678}{0.0050} = 13.56 \tag{53}$$

A t table can be found in Annex 5, and after consulting it, we find a critical value of 2.82 for 9 degrees of freedom and 99% confidence. Therefore, under this test, we can reject the null hypothesis and conclude that the accuracy resulting from testing for time-use satisfaction in academic activities using the WAIT-2022 initial survey dataset is greater than that of a random classifier with a uniform distribution.

The second test, at 99% confidence, with $n-1$ degrees of freedom, is performed with the following hypotheses in mind:

H$_0$: The accuracy from testing the WAIT-2022 initial survey dataset on a DT-based classifier trained with ENUT-2019 is equal to or less than the accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime. This hypothesis is expressed as follows:

$$m \geq 0.6035 \tag{54}$$

H$_1$: The accuracy from testing the WAIT-2022 initial survey dataset on a DT-based classifier trained with ENUT-2019 is greater than the average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime. This hypothesis is expressed as:

$$m < 0.6035 \tag{55}$$

As the directionally of the effect of interest is reversed, the calculations are as follows using the absolute value:

$$t = \left|\frac{0.5678 - 0.6035}{0.01579 / \sqrt{10}}\right| = \left|\frac{-0.0357}{0.0050}\right| = 7.14 \tag{56}$$

Again, after consulting the t table found in Annex 5, we find a critical value of 2.82 for 9 degrees of freedom and 99% confidence. Therefore, under this test, we can reject the null hypothesis and conclude that the accuracy resulting from testing for time-use satisfaction in academic activities using the WAIT-2022 initial survey dataset is greater than the average of the accuracy calculated under a 10-fold cross-validation regime using the ENUT-2019 preprocessed dataset as used throughout this dissertation.

While time-use satisfaction in academic activities is our variable of interest, and the measurement instrument that resulted in the WAIT-2022 initial survey was designed to predict this variable value in particular, the accuracy of the other time-use satisfaction variable is also relevant. The same tests as performed with time-use satisfaction in academic activities can be done for each time-use satisfaction variable; the data necessary for such tests is shown in **Table 48**, in which, for clarity, the values for time-use satisfaction in academic activities are also included.

Table 48: *Mean, standard deviation, sample size, and degrees of freedom for accuracy of time-use satisfaction contained in the ENUT-2019; and the accuracies using as test dataset the WAIT-2022 initial survey dataset in an RF classifier trained with ENUT-2019 data*

Time-use satisfaction in…	ENUT-2019			DF.	WAIT-2022	Critical value
	Mean	Std. Dev.	N		μ_2	
Domestic	0.6005	0.018675	10	9	0.4758	2.82
Academic activities	0.5678	0.01579	10	9	0.6035	2.82
Work	0.9177	0.010450	10	9	0.7621	2.82
Support to close ones	0.6781	0.018640	10	9	0.5947	2.82
Social/family	0.6707	0.012887	10	9	0.6696	2.82

Commutes	0.6432	0.016336	10	9	0.6872	2.82
Payments/paperwork	0.8997	0.031080	10	9	0.7533	2.82
Preferred activities	0.5716	0.010268	10	9	0.8194	2.82

Table 49 shows the results of t-tests using the data in **Table 48** performed with the variables linked to time-use satisfaction for the following hypotheses at 99% confidence, with $n-1$ degrees of freedom, with $\mu_1 = 0.5$ and μ_2 equal to the accuracies for time-use satisfaction for each kind of activity using the WAIT-2022 as test dataset.

The same hypotheses tested with time-use satisfaction in academic activities were tested for time-use satisfaction in all kinds of activities included in the ENUT and WAIT-2022 surveys. The first set of tests were performed with the following hypotheses in mind:

H$_0$: The average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime is equal to or less than that of a random classifier with a uniform distribution. This hypothesis is expressed as follows:

$$m \leq \mu_1 \tag{57}$$

H$_1$: The average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime is greater than that of a random classifier with a uniform distribution. This hypothesis is expressed as:

$$m > \mu_1 \tag{58}$$

The second set of tests, at 99% confidence, with $n-1$ degrees of freedom, are performed with the following hypotheses in mind:

H$_0$: The accuracy from testing the WAIT-2022 initial survey dataset on a DT-based classifier trained with ENUT-2019 is equal to or less than the average accuracy using a DT-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime. This hypothesis is expressed as follows:

$$m \geq \mu_2 \tag{59}$$

H$_1$: The accuracy from testing the WAIT-2022 initial survey dataset on a DT-based classifier trained with ENUT-2019 is greater than the average accuracy using a decision tree-based classifier trained with ENUT-2019 data under a 10-fold cross-validation regime. This hypothesis is expressed as:

$$m < \mu_2 \tag{60}$$

In the second test, the directionally of the effect of interest is reversed; in this case, the t values are multiplied by -1 and compared to the critical value as is.

***Table 49**: Results of Student's t-test for the accuracy of each time-use satisfaction variables contained in the ENUT-2019 and WAIT-2022 surveys for the main hypotheses and supplementary hypotheses*

Activities	**H_0: Worse than random classifier ($\mu_1 = 0.5$)**		**H_0: Worse than ENUT-2019 cross-validation regime (μ_2)**	
	t	Result	t*	Result
Domestic	16.138	Rejects H_0	-20.026	Fails to reject H_0
Academic activities	13.56	Rejects H_0	7.14	Rejects H_0
Work	119.901	Rejects H_0	-44.658	Fails to reject H_0
Support to close ones	28.662	Rejects H_0	-13.421	Fails to reject H_0
Social/family	39.738	Rejects H_0	-0.256	Fails to reject H_0
Commutes	26.303	Rejects H_0	8.075	Rejects H_0
Payments/paperwork	38.583	Rejects H_0	-14.133	Fails to reject H_0
Preferred activities	20.919	Rejects H_0	72.399	Rejects H_0

**The directionality of the test is reversed; therefore, the t values were multiplied by -1 and then compared as is with the critical value from the t table*

As can be seen in **Table 49**, the accuracies for all variables of time-use satisfaction under a 10-fold cross-validation regime using the ENUT-2019 dataset are statistically significantly better than a random classifier with a uniform distribution, validating the predictive power of the ENUT-2019 preprocessed dataset that is then used to train the WAIT-2022 decision tree-based classifier. On the other hand, only accuracy for time-use satisfaction in academic and preferred activities and commutes is better than the average accuracy from the cross-validation regime using the ENUT-2019 preprocessed dataset. This result means that only these three time-use satisfaction variables benefit from the AI-aided model selection performed to design the WAIT-2022 initial survey.

VIII.3. Recommendations given by WAIT-2022 to students

Having validated the accuracy of the intelligent algorithms used to build WAIT-2022, it is crucial to notice that the system can only give recommendations to enhance time-use satisfaction to users who it predicts do not have it, even if the users deem themselves as satisfied. Conversely, it can only give general advice to users who deem themselves unsatisfied but that the classifiers predict should have time-use satisfaction.

One of the rules of the study is that personalized recommendations were only given to users who identified themselves as unsatisfied and were accurately predicted so by WAIT-2022. With an accuracy of 0.6035 and 147 users indicating an absence of time-use satisfaction in academic activities, as shown in **Table 44**, the model should have about 88 opportunities to give proper recommendations depending on the run.

For the remaining students that WAIT-2022 was unable to produce specific recommendations to boost time-use satisfaction in academic activities, the recommendations were based in the generic results of a 400-representative sample of ENUT-2014 respondents who are students without time-use satisfaction in academic activities as classified by an RF classifier trained with the ENUT-2019 preprocessed dataset, a summary of the sets of the five most frequent recommendations to increase and decrease particular activities sent to their peers who did get sent personalized recommendations, and the generic recommendations offered by the Wellbeing Five model seen in **Table 1**.

The ENUT-2014 respondents who are students sample received an average of 1.025 recommendations. This means most respondents were issued a single recommendation. A summary of the percentages of the total number of recommendations to the ENUT-2014 respondents taken by each activity can be seen in **Table 50**.

Table 50: *Summary of the percentage of the total number of recommendations to ENUT-2014 respondents taken by each activity with no penalties or restrictions to changes in time allocations to particular activities*

To increase	%	To decrease	%
Cleaning the house's interior[a]	18.3%	Cooking or preparing drinks	79.5%
Personal hygiene and care	0.2%	Watching TV or video streaming	0.5%
Cleaning the house's interior	0.2%	Sleeping	0.2%
Commuting to university/school	0.2%	Folding clothes	0.2%

E-mail, chat, social networks	0.2%	Sports and workouts[b]	0.2%
All others	0.0%	All others	0.0%

[a] Activity only during weekends [b] activity during both weekdays and weekends

In practice the data shown in **Table 50**, means that WAIT-2022 users with no personalized recommendations received a message letting them know that 18.3% of a nationwide representative sample of students were recommended to make more time for cleaning their domestic environment during weekends, while 79.5% were recommended to decrease the time assigned to cooking and preparing drinks during weekdays.

The data shown in **Table 50** raises a few questions. What do cooking and cleaning have to do with time-use satisfaction in academic activities? Some plausible hypotheses are that these students allocate so much time to their academic activities that they do not have time to clean their domestic environment even during the weekends, and that the time allocated to cooking and preparing meals could be better put to use for more academic activities. Nonetheless, this raises the question of agency.

What if most of these ENUT-2014 respondents received these two recommendations and some of them could not make the suggested changes. Perhaps the GA could be run again with a restriction or penalty added to sets of recommendations including changes to these activities in order to let the GA search for other routes down the trained decision tree into presence of satisfaction in academic activities. These kind of penalties to undesirable changes can be added via the k_i variables in the GA model presented in section VI.7. This approach was applied to the personalized recommendations sent to WAIT-2022 users as can be seen in **Table 51**.

***Table 51**: Percentages of the total number of recommendations to WAIT-2022 users by each activity with a different set of restrictions on changes of time allocation, and average number of recommendations per user for each set of restrictions*

Restrictions:	**No restrictions**		**Cooking**		**Cooking, Sleep**		**Cooking, Sleep, Safety**		**Cooking, Sleep, Safety, Religious**	
Average num. of recommendations	1.000 per user		1.161 per user		1.036 per user		1.036 per user		9.500 per user	
Activities	Dec.	Inc.	Dec.	Inc.	Dec.	Inc.	Dec.	Inc.	Dec.	Inc.
Work	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	1.7%
Sleep and naps	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	2.6%	0.8%
Sleep and naps*	0.0%	0.0%	0.0%	55.4%	0.0%	51.7%	0.0%	0.0%	0.0%	0.0%
Meals	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	1.9%
Meals*	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	60.3%	0.0%	0.0%	0.0%
Hygiene, per. care	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	0.9%
Hygiene, per. care *	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	1.1%
Classes	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%	2.1%
Homework	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	1.7%
Commutes	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%	1.9%
Cooking	28.2%	71.8%	1.5%	0.0%	0.0%	0.0%	12.1%	0.0%	10.5%	0.0%
Meals and dishes	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	1.7%
Cleaning house	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.1%	2.3%
Cleaning house*	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	1.1%
Folding clothes	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	0.0%	0.0%	2.4%	2.1%
Securing belongings	0.0%	0.0%	9.2%	21.5%	17.2%	27.6%	0.0%	0.0%	0.0%	0.0%
Sports, workouts	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	1.5%
Watching video	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	2.8%
E-mail, social net.	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	0.8%
E-mail, social net.*	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	1.3%	2.6%
Reading	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	1.7%
Artistic, cultural	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.3%	1.1%
Artistic, cultural*	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	1.5%	1.3%
Civic, political	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	2.1%	1.9%
Civic, political *	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	2.4%
Social, visiting	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	1.5%
Social, visiting *	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.3%	1.7%
Volunteering	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	1.1%
Volunteering*	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.9%	0.8%
Religious activities	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	1.5%
Religious activities*	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	24.1%	0.0%	0.0%
Conversation	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	1.5%
Conversation *	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	0.0%	0.9%	2.6%
Sum	28.2%	71.8%	16.9%	83.1%	19.0%	81.0%	75.9%	24.1%	53.9%	46.1%

TESIS TESIS TESIS TESIS TESIS

As already mentioned, an advantage of using a GA is that penalties can be added to specific characteristics or “genes” of the solutions, for example the value of k_i variable linked to the i activity change that needs to be restricted can be made greater than 1 and relatively big if x_i does not equal zero for a particular solution to the problem and will be felt when evaluation the objective function of the algorithm shown in **equation (47)**. For example, given that letting the GA search for a solution without restrictions or penalties to suggesting changes to any activity results in it only finding recommendations to increase or decrease time allocated to cooking and preparing meals and drinks, adding a penalty to such particular recommendations lets the GA find other set of recommendations such as decreasing the time allocated to securing belongings and the house, and increasing it for sleeps and naps.

But what if a WAIT-2022 user could not change the time it allocates to sleep and naps. Then this activity can be added to the set of activities with penalties for recommendations of change. However, restricting this recommendation led to the GA to difficulties finding solutions, and relaxing the restriction led to it still recommending increasing time for sleep and naps. Only when adding restrictions to cooking, sleep and safety activities did the GA find recommendations that included decreasing time for meals and increasing it for religious activities during weekends. Finally, if we add religious activities to the set of activities with penalties for recommendations involving it the GA finds itself limited to suggesting much more complex solutions. Up to this point, most WAIT-2022 users with personalized recommendations received a single recommendation to either increase or decrease time allocated to a particular activity, but adding too many restrictions makes the number of average recommendations to each user increase to 9.5 and now the recommendations are much more diverse and particular to each WAIT-2022 as can be seen in the right-most columns of **Table 51**.

WAIT-2022 users who received personalized recommendations just were told what they were in the five restrictions cases shown in **Table 51**. For example, a randomly selected WAIT-2022 user received the following recommendations:

- Decrease time assigned to cooking and preparing meals and drinks.
    - If the above recommendation is impractical or not useful, then increase the time assigned to sleep and naps during weekends.

        - If the above recommendations are impractical or not useful, then decrease the time assigned to securing your house and belongings.
            - If the above recommendations are impractical or not useful, then decrease the time assigned to meals.
                - If the above recommendations are impractical or not useful then:
                    - Decrease time assigned to:
                        - Cleaning
                        - Folding clothes
                        - Watching TV and videos
                        - Religious activities
                        - E-mail, chat, and social media
                    - Increase time assigned to:
                        - Classes
                        - Homework
                        - Artistic and cultural activities
                        - Conversation and advice with close ones

On the other hand, users with generic recommendations just got told the percentages of the two most frequent changes recommended for each case:

- Consider that 28.2% of users were recommended to decrease time assigned to cooking and preparing meals and drinks while 71.8% were recommended to increase time assigned to this activity.
    - For those for whom the above recommendations are impractical or not useful, 9.2% received a recommendation to decrease time assigned to securing their house and belongings, and 55.4% received a recommendation to increase time assigned to sleep and naps during weekends.
        - For those for whom the above recommendations are impractical or not useful, the system still recommended more time for sleep and naps during

weekends to 51.7% and redoubled with a recommendation to decrease time assigned to securing their house and belongings for 17.2% of the users.

- When forced to find other main suggestions other than increasing sleep and decreasing time securing belongings, 60.3 received recommendations to decrease time assigned to meals and 24.1 to increase time assigned to religious activities during weekends.
    - For those for whom the above recommendations are impractical or not useful, the system developed more complex plans involving an average of 9 or more recommendations per user.

A disclaimer was added to all recommendations given by WAIT-2022, that while a link between the activities recommended and time-use satisfaction exists, it is not necessary causal, but perhaps people who have certain kinds of time use patterns are more likely to have time-use satisfaction. For example, people who do take more time to sleep or go to religious activities during weekends may do so because they have free time to it, not because the activity itself leads to time-use satisfaction, or perhaps activities such as good sleep and cooking meals do lead in some way to time-use satisfaction. Further research would be needed to test these hypotheses.

VIII.4. Correlation of agency-based wellbeing to SWB

Including the agency-based wellbeing variable for time use as an innovation presented in this dissertation is of ample interest for researching the interactions between time management, wellbeing, and performance. A first step towards validating an agency-based wellbeing construct is taken in the research performed with the implementation of WAIT-2022. A correlation analysis was done for each SWB variable included in the ENUT-2019 survey towards the agency-based wellbeing variable, that is, the variable linked to improving the capacity to interact with the environment and change it.

As mentioned, the correlation analysis uses the Pearson coefficient, which is recommended for this kind of data in the field literature (Ark, 2014). The results of the correlation analysis are shown in **Table 52**.

***Table 52:** Correlations of variable linked to wellbeing-agency to all other SWB variables and some sums of these (values above 0.2 are in bold)*

Variable of subjective wellbeing	Correlation to agency-based wellbeing
Time-use satisfaction sum	0.10717
Satisfaction in life in general	0.12438
Satisfaction in particular domains sum*	**0.22955**
Satisfaction in particular domains sum + life in general*	**0.22114**
Time-use satisfaction in academic activities	0.02714
Time-use satisfaction in work activities	0.01098
Time-use satisfaction in domestic activities	0.01326
Time-use satisfaction in preferred activities	0.00423
Time-use satisfaction in support of close ones	0.07485
Time-use satisfaction in social and family activities	-0.00715
Time-use satisfaction in commutes	0.11945
Time-use satisfaction in payments and paperwork	0.13961
Satisfaction in family life	0.12435
Satisfaction in affective life	0.17683
Satisfaction in social life	0.07757
Satisfaction in the personal economic situation	**0.24081**
Satisfaction with the housing situation	0.17297
Happiness	0.08453

**These are construct variables, which are the sum of the satisfaction values*

The results shown in **Table 52** show some promise in further developing an agency-based wellbeing construct connected to the well-researched SWB ones. Social science literature generally considers correlations between -0.2 and 0.2 as weak and between 0.2 and 0.5 as moderate and meaningful (Urdan, 2005). Such moderate correlations are marked in bold in **Table 52**. These correlations of the agency-based wellbeing variable exist with a robust construct of satisfaction in diverse domains of life and satisfaction with the personal economic situation. It should be noticed that these variables with a moderate correlation to the agency-based wellbeing variable are about satisfaction with a domain of life, not with time use in these domains. In fact, a robust time-use satisfaction construct would not fare as well in its correlation to agency-based wellbeing as a sum of all time-use satisfaction correlation is just above 0.1, driven mainly by time-use satisfaction in commutes and payments and paperwork.

IX. Conclusions and future research

In broad strokes, our main research question is answered in the positive for the specific case of our variable of interest for which the WAIT-2022 measurement instrument was designed, that of time-use satisfaction in academic activities, which is the SWB variable most directly linked to both time management and the sample of users which were all students. Consequently, we conclude that an AI-powered time management tool (in this case, WAIT-2022) can accurately predict SWB variables, of which time-use satisfaction is one, in particular, time-use satisfaction in academic activities for which the intelligent time management tool was specifically designed for, but also for time-use satisfaction in commutes and preferred activities. Other wellbeing variables may be affected too via their links to time-use satisfaction.

The conclusions of this dissertation can be expressed in a two-laned way:

1. An intelligent time management tool, developed with the aid of AI by using an automated model selection process, can conserve or enhance the accuracy of testing the training datasets in ML classifiers in a cross-validation regime, that is also better than the accuracy of a random classifier with a uniform distribution, for a variable or set of variables of interest.
2. The WAIT-2022 initial survey results show promise for further developing an agency-based wellbeing construct.

The main contribution of this dissertation is the conclusion that as intelligent time management tools based on interpretable ML algorithms can have significant accuracy, they can issue valid recommendations to the subset of users they correctly predict to need them if tested or used with labeled data. Hence, there is great potential in developing such tools and the practice of intelligent time management to realize the benefits of well-practiced time management further. Improving the accuracy levels of such tools would allow them to enlarge the set of users they can issue recommendations to; however, they can already be designed to avoid issuing recommendations to users for which they incorrectly predict their condition.

In the following sections of this chapter, we briefly develop the conclusions referent to other contributions of the research performed for this dissertation, as well as some closing thoughts and what we consider avenues for further research.

IX.1. WAIT-2022 project and a case for AI-aided design of measurement instruments

One of the most important accomplishments of the research presented in this dissertation is the successful AI-aided design of a measurement instrument in the form of the WAIT-2022 initial survey. This technique, although already presented in the research team's previous work (Marin et al., 2021), was applied to fruition during the construction of WAIT-2022 and tested to have not only conserved accuracy in its variable of interest but have a slightly enhanced level of it.

This result means that it is possible to use AI tools to automate a portion of the design of a smaller, more focused measurement instrument by employing automated model selection techniques using a dataset produced by a bigger, well-designed instrument design, such as a nationwide survey. In this case, the new measurement instrument has fewer variables than the original one, conserving only the most relevant ones to predict a variable of interest. This briefer and more focused instrument can be further validated using another dataset produced by the original measurement instrument, that is, using another sample or by testing it with new data. Both methods of validation were successfully implemented in this dissertation.

These conclusions are an important contribution to the application of AI to the fields of social sciences, as new, brief measurement instruments focused on predicting a variable of research interest can be designed in less time than from scratch, and the process can be mathematically validated. Also, this process does not preclude the further modification of new measurement instruments, so long as these modifications are part of a standard design process for measurement instruments (Marin et al., 2021) or use other parts of the original measurement instrument.

IX.2. The promises of an agency construct for time use

In the first parts of this dissertation, we not only presented a set of wellbeing definitions and constructs already well known in the wellbeing literature but also presented a definition of our own based on the insights of Spinoza and Aristotle that agency is directly linked to human wellbeing, hence an agency-based definition of wellbeing in which we posited that wellbeing is the state of having the capacities and abilities to interact with our environments while having these capacities and abilities increased to further interact with our environments and even change them.

A variable related to agency-based wellbeing linked to particular activities was included in the WAIT-2022 survey. It had a moderate correlation to the construct of satisfaction in different domains of life and in general. However, this correlation is driven mainly by the variable of satisfaction in the economic situation and, to a lesser degree, satisfaction in family life, affective life, and housing situations. Unfortunately, for this research, agency-based wellbeing has a low correlation to time-use satisfaction. This conclusion means that whether students perceive that their activities lead them to increase their capabilities to interact further with and change their environments has more to do with their judgment of their economic situation than with their judgments of the adequacy of the time allocated to their activities. While weak, even the correlation with satisfaction in affective, family, and housing situations is greater than with any time-use satisfaction variable.

Therefore, a moderate link is suggested between the cognitive evaluation of one's economic situation, of which housing is an important factor, and finding agency-based wellbeing in particular activities, with affective and family life coming close to a moderate correlation but remaining weaker than the link to the satisfaction with the economic situation.

IX.3. Answers to our research questions

As already surmised, the answer to our main research question, "Can intelligent time management tools trained with mostly time-use datasets accurately predict SWB variables?" is answered in the positive by the results and the statistical tests presented in the previous chapter.

As for the complementary research questions linked to other specific objectives:

1. Can basic principles of preserving and enhancing well-being be articulated to develop AI tools?
2. Can time-use survey data be characterized as human behavior in an economy of time?
3. Can interpretable machine learning algorithms trained mainly with time-use data be used for prediction and recommendations in wellbeing?

The whole of the present and previous chapter attests that the answer to the first two questions is yes: basic principles of preserving and enhancing well-being can be articulated to develop AI tools in such a way that they treat time-use survey data as variables that can represent human behavior (whether current or in the form of behavior change) in the context of the scarcity of time to assign to different activities just as in economic models there is always a scarce resource.

Finally, the accuracy validations presented in the previous chapter and the results in **Table 23** certainly let us answer in the positive the third question: Yes, interpretable machine learning algorithms trained mainly with time-use data can be used for prediction and recommendations for wellbeing enhancement even if only in some instances, in particular time-use satisfaction.

IX.4. Results that require further research

Further work is necessary for a longitudinal experiment analogous to our single survey experiment in which the members of the sample group give feedback about their SWB levels and their satisfaction with the use of the intelligent time management tool at different points in time after taking the initial survey and receiving the recommendations from the system. Even an interactive approach, in which the users select the restrictions on particular activities they do not want to change in terms of time allocated to them is on the table. While in the research presented in this dissertation we have shown that it is possible to build accurate intelligent time management tools that produce personalized recommendations to their users, research in the field of wellbeing and time management itself needs to be done to ascertain the effectiveness of such intelligent tools enhancing SWB levels across time and in comparison to traditional time management practices and tools. Only after this work has been done can the concept of intelligent time management be further developed empirically.

Other areas in need of further research based on the results presented in this dissertation are the further development of an agency-based definition of wellbeing and how a construct based on this definition can be developed. The leads are that the SWB variables most correlated to this definition are those linked to satisfaction, particularly satisfaction with the economic situation of the person, which would link agency to economic power, a result not surprising with the current social configurations within which college students live. Also, it would be of great interest for other researchers to create and test brief measurement instruments based on the datasets produced by bigger ones by using the techniques presented in this text.

Finally, while the replication of the experiment presented here with more extensive samples and better user feedback systems would improve on the results already presented if done with groups of students, new results would be produced if done with groups of members of organizations such as workers or employees which is the other group for which the WAIT-2022 initial survey measurement instrument has high performance in the prediction of time-use satisfaction. These newer results would shed light on the usefulness of intelligent time management tools such as WAIT-2022 in work and corporate environments in which regulations such as the NOM-035 and ESG social evaluations are relevant in order to mitigate risks related to the wellbeing of members of the organization, and to the organization as a whole.

The analysis of user data and the implementation of intelligent tools that can predict user attributes can produce interesting results that can lead to further research, such as the big difference in satisfaction towards the allocation of time to preferred activities in students in the WAIT-2022 sample of college students as compared with students from a broader nationwide representative sample. Perhaps different kinds of students have different judgments of how much time they allocate to their preferred activities, or maybe there is a factor impeding this SWB dimension of students at the university or groups in which WAIT-2022 was applied.

Other interesting result were the recommendations given to the ENUT-2014 respondents who are students that increasing time to clean one's domestic environment during weekends and decreasing time allocated to cook during the weekdays can advance most students towards time-use satisfaction in their academic activities. Experiments involving giving a group of students a free meal plan during the weekdays or reducing homework loads to free time for domestic activities, at least on weekends, can shed light on these issues. The research team expressed interest in finding out if this pattern held for the sample that took the WAIT-2022 initial survey and this hypothesized problem could be present in their university, however while the GA did find that 28.2% of students were recommended to decrease time assigned to cooking and preparing meals and drinks, the remaining 71.8% of students received the opposite recommendation and changes to the time allocated to cleaning were rarely recommended. Hence it would seem that the hypothesized problem is not as prevalent in the research team’s university, and perhaps time to cook is a mediator between time-use satisfaction in academic activities and other omitted variable, or a parallel effect of such variable. Finding out more about these questions would require further research.

IX.5. Closing thoughts

The study of human wellbeing is a practice-oriented multidisciplinary endeavor, just like ML and other fields of AI. In both endeavors, research, and practice live and die by the data available, the quantity and quality of which are of utmost importance to both statistical significance and algorithms' accuracy. Therefore, it is unsurprising that both activities share some obstacles and are complementary, particularly when using ML algorithms that are, in principle, interpretable by the human mind, such as tree-based classification, and not models such as ANNs that are predominantly black boxes.

One of the most essential contributions of this research is the promising use of AI to research and enhance human wellbeing not only in isolation but in conjunction with data and activities pertinent to people's daily lives. We argue that it is not the case that AI should be used to build specific tools to enhance wellbeing, but that wellbeing enhancement should be integrated into most AI tools that touch people's lives. Just as is the case with the research presented in this dissertation, in which wellbeing enhancement was integrated into an intelligent time management tool, a tool circumscribed in a discipline that is, in principle, interested only in furthering performance and optimizing the use of time, which affects SWB incidentally.

The combination of a wellbeing enhancement criterium and time management principles resulted in the introduction of the concept of intelligent time management that can, in principle, lead to the greatest performance advantages in comparison to other traditional time management practices, but for it not to lead people to their limits in an unhealthy way necessarily needs a wellbeing criterium or restriction integrated. We are curious whether that can be the case in many other disciplines and practices in which AI is currently integrated. Further wellbeing research is thus justified and needed in other fields where AI is applied and directly impacts people.

References

Aeon, B., & Aguinis, H. (2017). It's About Time: New Perspectives and Insights on Time Management. *Academy of Management Perspectives*, *31*(4), 309–330. https://doi.org/10.5465/amp.2016.0166

Aggarwal, C. C. (2018). An Introduction to Neural Networks. In *Neural Networks and Deep Learning: A Textbook* (pp. 1–52). Springer International Publishing. https://doi.org/10.1007/978-3-319-94463-0_1

Akerkar, R. (2019). Deep Learning. In *Artificial Intelligence for Business* (pp. 33–40). Springer International Publishing. https://doi.org/10.1007/978-3-319-97436-1_3

Althbiti, A., & Ma, X. (2020). Machine Learning. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data* (pp. 1–5). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_539-1

Aristotle. (2013). Book I: Slavery and the Will to Power. In C. Lord (Ed.), *Aristotle's "Politics"* (2nd ed., pp. 1–24). The University of Chicago Press.

Ark, N. (2014). Zero-Order Relationships. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 7311–7313). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_3306

Axtell, C., Taylor, M., & Wessels, B. (2019). Big data and employee wellbeing: Walking the tightrope between Utopia and Dystopia. *Social Sciences*, *8*(12), 1–5. https://doi.org/10.3390/socsci8120321

Ayyadevara, V. K. (2018). Artificial Neural Network. In *Pro Machine Learning Algorithms : A Hands-On Approach to Implementing Algorithms in Python and R* (pp. 135–165). Apress. https://doi.org/10.1007/978-1-4842-3564-5_7

Ba, L. J., & Caruana, R. (2014). Do deep nets really need to be deep? *Advances in Neural Information Processing Systems*, *3*(January), 2654–2662.

Baid, V., & Jayaraman, V. (2022). Amplifying and promoting the "S" in ESG investing: the case for social responsibility in supply chain financing. *Managerial Finance*, *48*(8), 1279–1297. https://doi.org/10.1108/MF-12-2021-0588

Bainbridge, M. (2019). Big Data Challenges for Clinical and Precision Medicine. In M. Househ, A. W. Kushniruk, & E. M. Borycki (Eds.), *Big Data, Big Challenges: A Healthcare Perspective: Background, Issues, Solutions and Research Directions* (pp. 17–31). Springer International Publishing. https://doi.org/10.1007/978-3-030-06109-8_2

Bambrick, N. (2016). *Support Vector Machines: A Simple Explanation*. KD Nuggets. https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html

Batarseh, F. A. (2018). Artificial Intelligence. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_225-1

Becchetti, L., Bobbio, E., Prizia, F., & Semplici, L. (2022). Going Deeper into the S of ESG: A Relational Approach to the Definition of Social Responsibility. *Sustainability*, *14*(15). https://doi.org/10.3390/su14159668

Bellet, C., & Frijters, P. (2019). Big Data and Well-being. In J. F. Helliwell, R. Layard, & J. D. Sachs (Eds.), *World Happiness Report* (pp. 97–122). Sustainable Development Solutions Network.

Bijlstra, G., & Dotsch, D. (2012). *FaceReader 4 - RaFD* (Issue 1).

Blanco-Justicia, A., & Domingo-Ferrer, J. (2019). Machine Learning Explainability Through Comprehensible Decision Trees. In P. and T. A. M. and W. E. Holzinger Andreas and Kieseberg (Ed.), *Machine Learning and Knowledge Extraction* (pp. 15–26). Springer International Publishing.

Boden, M. A. (2014). GOFAI. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 89–107). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139046855.007

Boehm, J. K., Chen, Y., Qureshi, F., Soo, J., Umukoro, P., Hernandez, R., Lloyd-Jones, D., & Kubzansky, L. D. (2020). Positive emotions and favorable cardiovascular health: A 20-year longitudinal study. *Preventive Medicine*, *136*(March), 106103. https://doi.org/10.1016/j.ypmed.2020.106103

Boniwell, I. (2006). *Satisfaction with time use and its relationship with subjective well-being* [PhD thesis, The Open University]. https://doi.org/10.21954/ou.ro.0000e971

Bradley, B. (2021). Give Me an 'S'! Investigating the Social Aspects of ESG. In *ESG Investing For Dummies* (1st ed., pp. 63–82). John Wiley & Sons, Inc.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. In *Classification and Regression Trees* (1st ed.). Chapman & Hall/CRC Press. https://doi.org/10.1201/9781315139470

Bryson, A., Forth, J., & Stokes, L. (2017). Does employees' subjective well-being affect workplace performance? *Human Relations*, *70*(8), 1017–1037. https://doi.org/10.1177/0018726717693073

Bücker, S., Nuraydin, S., Simonsmeier, B. A., Schneider, M., & Luhmann, M. (2018). Subjective well-being and academic achievement: A meta-analysis. *Journal of Research in Personality*, *74*, 83–94. https://doi.org/10.1016/j.jrp.2018.02.007

Capterra. (2019a). *Best Customer Satisfaction Software 2020*. Reviews of the Most Popular Tools & Systems. https://www.capterra.com/customer-satisfaction-software/

Capterra. (2019b). *Best Employee Monitoring Software 2020*. Reviews of the Most Popular Tools & Systems. https://www.capterra.com/employee-monitoring-software/

Capterra. (2020). *Best Employee Engagement Software 2020*. Reviews of the Most Popular Tools & Systems. https://www.capterra.com/employee-engagement-software/

Caruana, R., & Niculescu-Mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, 161–168. https://doi.org/10.1145/1143844.1143865

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011). Evaluating Automatic Model Selection. *Journal of Time Series Econometrics*, *3*(1). https://doi.org/10.2202/1941-1928.1097

Charzyńska, E., Sussman, S., & Atroszko, P. A. (2021). Profiles of potential behavioral addictions' severity and their associations with gender, personality, and well-being: A person-centered approach. *Addictive Behaviors*, *119*. https://doi.org/10.1016/j.addbeh.2021.106941

Chen, L. (2018). *Support Vector Machines Simply Explained*. Towards Data Science. https://towardsdatascience.com/support-vector-machine-simply-explained-fee28eba5496

Chipman, H. A. (2011). Recursive Partitioning. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1202–1206). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_483

Chollet, F. (2018). Fundamentals of Deep Learning. In T. Arritola (Ed.), *Deep Learnig with Python* (1st ed., pp. 1–115). Manning Publications Co.

Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L. A. (2007a). Feature Extraction and Selection Methods. In *Data Mining: A Knowledge Discovery Approach* (pp. 133–33). Springer US. https://doi.org/10.1007/978-0-387-36795-8_1

Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L. A. (2007b). The Knowledge Discovery Process. In *Data Mining: A Knowledge Discovery Approach* (pp. 9–24). Springer US. https://doi.org/10.1007/978-0-387-36795-8_2

Claessens, B. J. C., van Eerde, W., Rutte, C. G., & Roe, R. A. (2007). A review of the time management literature. *Personnel Review*, *36*(2), 255–276. https://doi.org/10.1108/00483480710726136

Clark, A. E., Diener, E., Georgellis, Y., & Lucas, R. E. (2008). Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis. *The Economic Journal*, *118*(529), F222–F243. https://doi.org/10.1111/j.1468-0297.2008.02150.x

Collins, S., Sun, Y., Kosinski, M., Stillwell, D., & Markuzon, N. (2015). Are You Satisfied with Life?: Predicting Satisfaction with Life from Facebook. In N. Agarwal, K. Xu, & N. Osgood (Eds.), *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 24–33). Springer International Publishing.

Committee on National Statistics National Research Council. (2000). Survey Design Issues. In M. Ver Ploeg, J. Altonji, N. Bradburn, J. DaVanzo, W. Nordhaus, & F. Samaniego (Eds.), *Time-Use Measurement and Research: Report of a Workshop* (pp. 36–49). National Academy Press.

Cowling, P. I., Greenberg, H. J., & De Jong, K. (2013). Artificial Intelligence. In S. I. Gass & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (pp. 78–83). Springer US. https://doi.org/10.1007/978-1-4419-1153-7_42

Crisp, R. (2017). Well-Being. In E. N. Zalta (Ed.), *The {Stanford} Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.

Cristianini, N., & Ricci, E. (2016). Support Vector Machines. In M.-Y. Kao (Ed.), *Encyclopedia of Algorithms* (pp. 2170–2174). Springer New York. https://doi.org/10.1007/978-1-4939-2864-4_415

Cummins, R. A. (2013). Subjective Well-Being, Homeostatically Protected Mood and Depression: A Synthesis. In A. Delle Fave (Ed.), *The Exploration of Happiness: Present and Future Perspectives* (pp. 77–95). Springer Netherlands. https://doi.org/10.1007/978-94-007-5702-8_5

Das, A. (2014). Logistic Regression. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3680–3682). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1689

Denollet, J., Vaes, J., & Brutsaert, D. L. (2000). Inadequate response to treatment in coronary heart disease: adverse effects of type D personality and younger age on 5-year prognosis and quality of life. *Circulation*, *102*(6), 630–635.

Detel, W. (2001). Social Constructivism. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14264–14267). Pergamon. https://doi.org/https://doi.org/10.1016/B0-08-043076-7/01086-X

DeVoe, S. E., & Pfeffer, J. (2007). Hourly Payment and Volunteering: The Effect of Organizational Practices on Decisions about Time Use. *The Academy of Management Journal*, *50*(4), 783–798. https://doi.org/10.2307/20159890

Diener, E. (1984). Subjective Well-Being. *Psychological Bulletin*, *95*(3), 542–575. https://doi.org/10.1037/0033-2909.95.3.542

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, *55*(1), 34–43. https://doi.org/10.1037/0003-066X.55.1.34

Diener, E., Pressman, S. D., Hunter, J., & Delgadillo-Chase, D. (2017). If, Why, and When Subjective Well-Being Influences Health, and Future Needed Research. *Applied Psychology: Health and Well-Being*, *9*(2), 133–167. https://doi.org/https://doi.org/10.1111/aphw.12090

Diener, E., & Ryan, K. (2008). *Subjective well-being : a general overview*. *39*(4), 391–406.

Diener, E., & Ryan, K. (2009). Subjective Well-Being: A General Overview. *South African Journal of Psychology*, *39*(4), 391–406. https://doi.org/10.1177/008124630903900402

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: three decades of progress. In *Psychological Bulletin* (Vol. 125, pp. 276–302). http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.125.2.276

Diez, D., Cetinkaya-Rundel, M., & Barr, C. (2019). *OpenIntro Statistics* (4th ed.). OpenIntro.

Dobra, A. (2009a). Decision Tree Classification. In L. LIU & M. T. ÖZSU (Eds.), *Encyclopedia of Database Systems* (pp. 765–769). Springer US. https://doi.org/10.1007/978-0-387-39940-9_554

Dobra, A. (2009b). Decision Tree Classification. In M. T. LIU LING and ÖZSU (Ed.), *Encyclopedia of Database Systems* (pp. 765–769). Springer US. https://doi.org/10.1007/978-0-387-39940-9_554

Dua, D., & Graff, C. (2017). UCI Machine Learning Repository. In *University of California, Irvine, School of Information and Computer Sciences*.

Durand, M. (2015). The OECD better life initiative: How’s life? And the measurement of well-being. *Review of Income and Wealth*, *61*(1), 4–17. https://doi.org/10.1111/roiw.12156

Emmons, R. A., & Diener, E. (1985). Personality Correlates of Subjective Well-Being. *Personality and Social Psychology Bulletin*, *11*(1), 89–97. https://doi.org/10.1177/0146167285111008

Eurostat. (2019). *Time use survey (tus)*. Eurostat. https://ec.europa.eu/eurostat/cache/metadata/en/tus_esms.htm

Fayyad, U. M. (1997). Editorial. *Data Mining and Knowledge Discovery*, *1*(1), 5–10. https://doi.org/10.1023/A:1009715820935

Fürnkranz, J. (2010). Decision Tree. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 263–267). Springer US. https://doi.org/10.1007/978-0-387-30164-8_204

Fürnkranz, J. (2017). Decision Tree. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 330–335). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_66

Gaier, A., & Ha, D. (2019). Weight agnostic neural networks. *Advances in Neural Information Processing Systems*, *32*(NeurIPS), 1–19.

Gargiulo, R. A., & Stokes, M. A. (2009). Subjective well-being as an indicator for clinical depression. *Social Indicators Research*, *92*(3), 517–527. https://doi.org/10.1007/s11205-008-9301-0

Gass, S. I., & Fu, M. C. (Eds.). (2013). Machine Learning. In *Encyclopedia of Operations Research and Management Science* (p. 909). Springer US. https://doi.org/10.1007/978-1-4419-1153-7_200425

Gershuny, J. (2015). Time Use Research Methods. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (Second Edi, Vol. 24). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.44058-4

Goode, L. (2020, April 14). *Can a Wearable Detect Covid-19 Before Symptoms Appear?* WIRED. https://www.wired.com/story/wearable-covid-19-symptoms-research/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep feedforward networks. In *Deep Learning*. MIT Press. https://www.deeplearningbook.org/contents/mlp.html

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Google. (2022, July 18). *Classification: True vs. False and Positive vs. Negative*. Google for Developers. https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

Google. (2023, November 14). *Machine Learning Glossary*. Google for Developers. https://developers.google.com/machine-learning/glossary

Granqvist, N., & Gustafsson, R. (2016). Temporal institutional work. *Academy of Management Journal*, *59*(3), 1009–1035. https://doi.org/10.5465/amj.2013.0416

Gregory, R. L. (2006). *Knowledge by acquaintance, and knowledge by description*. Oxford University Press. https://doi.org/10.1093/acref/9780198662242.013.0499

Grills, S. (2017). Social Construction. In D. C. Poff & A. C. Michalos (Eds.), *Encyclopedia of Business and Professional Ethics* (pp. 1–4). Springer International Publishing. https://doi.org/10.1007/978-3-319-23514-1_29-1

Hadi, A. S. (2011). Expert Systems. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 480–482). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_240

Harrel, F. (1996). *Problems with stepwise regression*. Stata | FAQ. https://www.stata.com/support/faqs/statistics/stepwise-regression-problems/

Harvey, A. S., & Pentland, W. E. (2002). Time Use Research. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time Use Research in the Social Sciences* (pp. 3–18). Springer US. https://doi.org/10.1007/0-306-47155-8_1

Harvey, F. (2020, May 10). *Britons want quality of life indicators to take priority over economy | Health | The Guardian*. The Guardian. https://www.theguardian.com/society/2020/may/10/britons-want-quality-of-life-indicators-priority-over-economy-coronavirus

Health IQ. (2020). *Mission*. Health IQ Special Rate Insurance. https://www.healthiq.com/mission

Healthways. (2015). *Gallup-Healthways Well-Being 5 Fact Sheet*.

Hedman, E., Miller, L., Schoen, S., Nielsen, D., Goodwin, M., & Picard, R. (2012). Measuring autonomic arousal during therapy. *8th International Conference on Design and Emotion: Out of Control - Proceedings*.

Herrin, J., Witters, D., Roy, B., Riley, C., Liu, D., & Krumholz, H. M. (2018). Population well-being and electoral shifts. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0193401

Holland, J. H. (1975). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. In *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* U Michigan Press.

TESIS TESIS TESIS TESIS TESIS

Holsapple, C. W., & Whinston, A. B. (2013). Expert Systems. In S. I. Gass & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (pp. 528–532). Springer US. https://doi.org/10.1007/978-1-4419-1153-7_313

Hurka, T. (1993). *Perfectionism* (1st ed.). Oxford University Press.

INEGI. (2015). *Documento Metodológico: Encuesta Nacional sobre Uso del Tiempo 2014*.

INEGI. (2020). *Nota Técnica: Encuesta Nacional Sobre Uso del Tiempo (ENUT) 2019*. https://www.inegi.org.mx/img/ico/ico_pdf.png

INEGI. (2022). *Nota Técnica: Encuesta Nacional de Bienestar Autorreportado 2021*. https://www.inegi.org.mx/contenidos/programas/enbiare/2021/doc/enbiare_2021_nota_tecnica.pdf

Intel Corporation. (2016). *Improve Employee Wellness through Predictive Analytics*.

Ironmonger, D. (2018). Time Use. In *The New Palgrave Dictionary of Economics* (pp. 13676–13681). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-349-95189-5_2109

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Joshi, A. V. (2020). Decision Trees. In *Machine Learning and Artificial Intelligence* (pp. 53–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-26622-6_6

Karacaoglu, G., Krawczyk, J. B., & King, A. (2019). Introduction and Overview. In *Intergenerational Wellbeing and Public Policy: An Integrated Environmental, Social, and Economic Framework* (pp. 3–26). Springer Singapore. https://doi.org/10.1007/978-981-13-6104-3_1

Kleene, S. C. (1956). *Representation of Events in Nerve Nets and Finite Automata* (pp. 3–42). Princeton University Press. https://doi.org/https://doi.org/10.1515/9781400882618-002

Kleinbaum, D. G., & Klein, M. (2010). Introduction to Logistic Regression. In *Logistic Regression: A Self-Learning Text* (pp. 1–39). Springer New York. https://doi.org/10.1007/978-1-4419-1742-3_1

Koç, P. (2017). Internet Addiction and Subjective Wellbeing in University Students. In *Article Journal of Positive Psychology & Wellbeing* (Vol. 2017, Issue 1). www.journalppw.com

Kristoffersen, I. (2018). Great expectations: Education and subjective wellbeing. *Journal of Economic Psychology*, *66*, 64–78. https://doi.org/10.1016/J.JOEP.2018.04.005

Leeth, C. (2011). Affect. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behavior and Development* (p. 55). Springer US. https://doi.org/10.1007/978-0-387-79061-9_69

Lempitsky, V. (2019). Autoencoder. In *Computer Vision: A Reference Guide* (pp. 1–6). Springer International Publishing. https://doi.org/10.1007/978-3-030-03243-2_862-1

Li, X., Dunn, J., Salins, D., Zhou, G., Zhou, W., Schüssler-Fiorenza Rose, S. M., Perelman, D., Colbert, E., Runge, R., Rego, S., Sonecha, R., Datta, S., McLaughlin, T., & Snyder, M. P. (2017). Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. *PLoS Biology*, *15*(1). https://doi.org/10.1371/journal.pbio.2001402

Livshin, I. (2019). Learning About Neural Networks. In *Artificial Neural Networks with Java: Tools for Building Neural Network Applications* (pp. 1–5). Apress. https://doi.org/10.1007/978-1-4842-4421-0_1

Logistic Regression. (2008). In *The Concise Encyclopedia of Statistics* (pp. 319–320). Springer New York. https://doi.org/10.1007/978-0-387-32833-1_237

Look, B. C. (2007). Perfection , Power and the Passions in Spinoza and Leibniz. *Revue Roumaine de La Philosophie*, *51*(1–2), 21–38.

Lorenz, O. (2018). Does commuting matter to subjective well-being? *Journal of Transport Geography*, *66*, 180–199. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2017.11.019

Macan, T. H. (1994). Time Management: Test of a Process Model. *Journal of Applied Psychology*, *79*(3), 381–391. https://doi.org/10.1037/0021-9010.79.3.381

Malkoc, S. A., & Tonietto, G. N. (2019). Activity versus outcome maximization in time management. *Current Opinion in Psychology*, *26*, 49–53. https://doi.org/10.1016/j.copsyc.2018.04.017

Man, K. F., Tang, K. S., & Kwong, S. (1999a). Introduction, Background and Biological Inspiration. In *Genetic Algorithms: Concepts and Designs* (1st ed., pp. 1–22). Springer-Verlag London Berlin Heidelberg.

Man, K. F., Tang, K. S., & Kwong, S. (1999b). Modifications to Genetic Algorithms. In *Genetic Algorithms: Concepts and Designs* (1st ed., pp. 23–44). Springer-Verlag London Berlin Heidelberg.

Marin, M. E., & Ponce, J. C. (2020). Intelligent Time Use Suggestions for Wellbeing Enhancement. *Research in Computer Science*, *149*(10), 101–114. https://www.rcs.cic.ipn.mx/rcs/2020_149_10/Intelligent%20Time%20use%20Suggestions%20for%20Wellbeing%20Enhancement.pdf

Marin, M. E., Ponce, J. C., Muñoz, Á. E., & Padilla, A. (2021). Peligros y ventajas del diseño de instrumentos de medición auxiliado con estrategias de selección de atributos automáticas. *Research in Computing Science*, *150*(5).

Martínez Rizo, F. (2010). Los indicadores como herramientas para la evaluación de la calidad de los sistemas educativos. *Sinéctica*, *35*, 1–17. https://doi.org/2007-7033

Meredith, S. (2022, December 25). *Well-being: These countries are looking beyond GDP and economic growth*. CNBC. https://www.cnbc.com/2022/12/26/well-being-these-countries-are-looking-beyond-gdp-and-economic-growth.html

Moller, A., & Deci, E. (2014). Intrinsic Motivation. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3378–3381). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1532

Montgomery, D. C. (2020). Introduction. In *Design and analysis of experiments* (10th ed., pp. 1–20). John Wiley and Sons.

Moreno, A., & Ruiz-Mirazo, K. (2015). Homeostasis. In M. Gargaud, W. M. Irvine, R. Amils, H. J. (Jim) Cleaves, D. L. Pinti, J. C. Quintanilla, D. Rouan, T. Spohn, S. Tirard, & M. Viso (Eds.), *Encyclopedia of Astrobiology* (pp. 1113–1114). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-44185-5_730

Munro, P. (2010). Backpropagation. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (p. 73). Springer US. https://doi.org/10.1007/978-0-387-30164-8_51

Murphy, K. P. (2012). Online learning and stochastic optimization. In *Machine learning: a probabilistic perspective* (1st ed., pp. 261–266). The MIT Press.

Nandy, A., & Biswas, M. (2018). Reinforcement Learning Basics. In *Reinforcement Learning : With Open AI, TensorFlow and Keras Using Python* (pp. 1–18). Apress. https://doi.org/10.1007/978-1-4842-3285-9_1

Nash, J. (2020). Affect. In D. A. Leeming (Ed.), *Encyclopedia of Psychology and Religion* (pp. 20–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-24348-7_12

Nielsen, M. A. (2015). Why are deep neural networks hard to train? In *Neural Networks and Deep Learning*. Determination Press.

Nilsson, N. J. (1997a). Neural Networks. In *Introduction to Machine Learning: An Early Draft of a Proposed Textbook* (pp. 39–68). Department of Computer Science, Stanford University. https://ai.stanford.edu/~nilsson/MLBOOK.pdf

Nilsson, N. J. (1997b). Preliminaries. In *Introduction to Machine Learning: An Early Draft of a Proposed Textbook* (pp. 1–15). Department of Computer Science, Stanford University. https://ai.stanford.edu/~nilsson/MLBOOK.pdf

Norma Oficial Mexicana NOM-035-STPS-2018, (2018).

Nozal, A. L. (2019). *The Economy of Well-being: Creating Opportunities for People's Well-being and Economic Growth*. www.oecd.org/sdd/publicationsdocuments/workingpapers/

Nussbaum, M. C. (2001). Emotions as judgments of value. In *Upheavals of thought: the intelligence of emotions* (pp. 19–85). Cambridge University Press.

OECD. (2011a). How's life?: Measuring well-being. In *How's Life?: Measuring Well-Being* (Vol. 9789264121). https://doi.org/10.1787/9789264121164-en

OECD. (2011b). Why does subjective well-being matter for well-being? In *How's Life?: Measuring well-being* (pp. 266–267). OECD Publishing. dx.doi.org/10.1787/9789264121164-en

OECD. (2019). *What's the Better Life Index?* OECD Better Life Index. http://www.oecdbetterlifeindex.org/about/better-life-initiative/

OECD. (2020). How's Life? in OECD countries. In *How's Life? 2020* (1st ed., pp. 18–61). OECD Publishing.

Okun, M. A., Stock, W. A., Haring, M. J., & Witter, R. A. (1984). Health and subjective well-being: A meta-analysis. *International Journal of Aging and Human Development*, *19*(2), 111–132. https://doi.org/10.2190/QGJN-0N81-5957-HAQD

Origel-Rivas, G., Rendón-Lara, E., María Abundez-Barrera, I., & Alejo-Eleuterio, R. (2020). Redes neuronales artificiales y árboles de decisión para la clasificación con datos categóricos. *Research in Computing Science*, *149*(8), 541–554.

Orpen, C. (1994). The effect of time-management training on employee attitudes and behavior: A field experiment. *Journal of Psychology: Interdisciplinary and Applied*, *128*(4), 393–396. https://doi.org/10.1080/00223980.1994.9712743

Park, N., & Peterson, C. (2019). It's not just the economy: The psychological well-being of an electorate also matters for election outcomes. *Journal of Positive Psychology*, *14*(2), 127–132. https://doi.org/10.1080/17439760.2017.1350743

Peeters, M. A. G., & Rutte, C. G. (2005). Time management behavior as a moderator for the job demand-control interaction. *Journal of Occupational Health Psychology*, *10*(1), 64–75. https://doi.org/10.1037/1076-8998.10.1.64

Perlow, L. A., Mazmanian, M., & Hansen, E. (2016). Unlocking the temporal paradox: Creating temporal control in an era of the ideal worker. *Making Time for Time: Extending Theories of Temporality in and across Organizations*.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

Rath, T., & Harter, J. K. (2010). *Wellbeing: The five essential elements* (1st ed.). Gallup Press.

Ray, S. (2015). *Understanding Support Vector Machines Algorithm*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

Rebala, G., Ravi, A., & Churiwala, S. (2019). Machine Learning Definition and Basics. In *An Introduction to Machine Learning* (pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-15729-6_1

Roskam, E. E. (1989). Operationalization, a superfluous concept. *Quality and Quantity*, *23*(3–4), 237–275. https://doi.org/10.1007/BF00172446

Salgado, J. F., Blanco, S., & Moscoso, S. (2019). Subjective well-being and job performance: Testing of a suppressor effect. *Revista de Psicologia Del Trabajo y de Las Organizaciones*, *35*(2), 93–102. https://doi.org/10.5093/jwop2019a9

Sammut, C. (2017). Genetic and Evolutionary Algorithms. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 566–567). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_334

Sammut, C., & Webb, G. I. (Eds.). (2010). Cross-Validation. In *Encyclopedia of Machine Learning* (p. 249). Springer US. https://doi.org/10.1007/978-0-387-30164-8_190

Sammut, C., & Webb, G. I. (Eds.). (2017a). Artificial Neural Networks. In *Encyclopedia of Machine Learning and Data Mining* (pp. 65–66). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_921

Sammut, C., & Webb, G. I. (Eds.). (2017b). Logistic Regression. In *Encyclopedia of Machine Learning and Data Mining* (pp. 780–781). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_951

Sammut, C., & Webb, G. I. (Eds.). (2017c). Model Selection. In *Encyclopedia of Machine Learning and Data Mining* (p. 845). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_551

Sammut, C., & Webb, G. I. (Eds.). (2017d). Random Forests. In *Encyclopedia of Machine Learning and Data Mining* (p. 1054). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_695

Sammut, C., & Webb, G. I. (Eds.). (2017e). Random Subspace Method. In *Encyclopedia of Machine Learning and Data Mining* (p. 1055). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_696

Sano, A. (2016). *Measuring college students' sleep, stress, mental health and wellbeing with wearable sensors and mobile phones* [Massachusetts Institute of Technology]. https://affect.media.mit.edu/pdfs/15.Sano-Thesis.pdf

Schlosser, M. (2019, October 28). *Agency*. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/agency/

scikit-learn developers. (2020). *sklearn.feature_selection.chi2 — .* Scikit-Learn 0.24.1 Documentation. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html#sklearn.feature_selection.chi2

Shalev-Shwartz, S., & Ben-David, S. (2013). Understanding machine learning: From theory to algorithms. In *Understanding Machine Learning: From Theory to Algorithms* (Vol. 9781107057). https://doi.org/10.1017/CBO9781107298019

Şimşek, Ö. F. (2009). Happiness revisited: Ontological well-being as a theory-based construct of subjective well-being. In *Journal of Happiness Studies* (Vol. 10, Issue 5, pp. 505–522). https://doi.org/10.1007/s10902-008-9105-6

Sivanandam S.N., & Deepa, S. N. (2008). Terminologies and Operators of GA. In *Introduction to Genetic Algorithms* (pp. 39–81). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73190-0_3

Sivazlian, B. D. (2009). Operations researchOperations Research. In C. A. Floudas & P. M. Pardalos (Eds.), *Encyclopedia of Optimization* (pp. 2694–2696). Springer US. https://doi.org/10.1007/978-0-387-74759-0_465

Skyrius, R., Giri\=unien\.e, G. e, Katin, I., Kazimianec, M., & Žilinskas, R. (2018). The Potential of Big Data in Banking. In S. Srinivasan (Ed.), *Guide to Big Data Applications* (pp. 451–486). Springer International Publishing. https://doi.org/10.1007/978-3-319-53817-4_17

Smith, G. (2018). Step away from stepwise. *Journal of Big Data*, *5*(1). https://doi.org/10.1186/s40537-018-0143-6

Sorensen, G., Sparer, E., Williams, J. A. R., Gundersen, D., Boden, L. I., Dennerlein, J. T., Hashimoto, D., Katz, J. N., McLellan, D. L., Okechukwu, C. A., Pronk, N. P., Revette, A., & Wagner, G. R. (2018). Measuring Best Practices for Workplace Safety, Health, and Well-Being: The Workplace Integrated Safety and Health Assessment. *Journal of Occupational and Environmental Medicine*, *60*(5), 430–439. https://doi.org/10.1097/JOM.0000000000001286

Spinoza, B. (1996). *Ethics* (M. E. Curley & S. Hampshire, Eds.). Penguin Books.

Steptoe, A., Deaton, A., & Stone, A. A. (2015). Subjective wellbeing, health, and ageing. *The Lancet*, *385*(9968), 640–648. https://doi.org/10.1016/S0140-6736(13)61489-0

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, *103*(2684), 677 LP – 680. https://doi.org/10.1126/science.103.2684.677

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. www.stiglitz-sen-fitoussi.fr

Torgo, L. (2017). Regression Trees. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 1080–1083). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_717

Trudel-Fitzgerald, C., Qureshi, F., Appleton, A. A., & Kubzansky, L. D. (2017). A healthy mix of emotions: underlying biological pathways linking emotions to physical health. *Current Opinion in Behavioral Sciences*, *15*, 16–21. https://doi.org/10.1016/j.cobeha.2017.05.003

United Nations Development Programme. (2019). *Human Development Index (HDI)*. Human Development Reports. http://hdr.undp.org/en/content/human-development-index-hdi

Urdan, T. C. (2005). Correlation. In D. Riegert (Ed.), *Statistics in Plain English* (2nd ed., pp. 75–88). Lawrence Erlbaum Associates.

Vogt, P. W. (2005). *Post Hoc Theorizing*. Dictionary of Statistics & Methodology. https://doi.org/10.4135/9781412983907

Vogt, P. W., Gardner, D. C., & Haeffele, L. M. (2012). General Introduction, Design, Sampling, and Ethics. In *When to use what research design* (pp. 1–7). The Guilford Press.

Wang, S., Tang, J., & Liu, H. (2017). Feature Selection. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 503–511). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_101

Waraczynski, M. (2009). Emotion. In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 1088–1092). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_2972

Warr, P., & Nielsen, K. (2018). Wellbeing and work performance. *Handbook of Well-Being*, *February*, 1–22. https://doi.org/nobascholar.com

Weinberg, A. I., & Last, M. (2019). Selecting a representative decision tree from an ensemble of decision-tree models for fast big data classification. *Journal of Big Data*, *6*(1), 23. https://doi.org/10.1186/s40537-019-0186-3

Wellbeing Economy Alliance. (2021, March 31). *What's Happening with the Wellbeing Economy Governments?* Wellbeing Economy Governments News. https://weall.org/whats-happening-with-the-wellbeing-economy-governments

World Health Organization. (2021). *Geneva Charter for Well-Being*.

World Health Organization. (2023). *Promoting well-being*. WHO Activities. https://www.who.int/activities/promoting-well-being

Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2013). Support Vector Machines. In *Robust Data Mining* (pp. 35–48). Springer New York. https://doi.org/10.1007/978-1-4419-9878-1_5

Zhou, H. (2020). Decision Trees. In *Learn Data Mining Through Excel: A Step-by-Step Approach for Understanding Machine Learning Methods* (pp. 125–148). Apress. https://doi.org/10.1007/978-1-4842-5982-5_9

Zulkifli, H. (2018, January 20). *Understanding Learning Rates and How It Improves Performance in Deep Learning | by* . Towards Data Science. https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-learning-d0d4059c1c10

Zuzanek, J., & Zuzanek, T. (2015). Of Happiness and of Despair, Is There a Measure? Time Use and Subjective Well-being. *Journal of Happiness Studies*, *16*(4), 839–856. https://doi.org/10.1007/s10902-014-9536-1

Annex 1: Survey text, original text in Spanish

The survey questionnaire is described first in the original version presented to the students, and then a translation to English.

Encuesta sobre bienestar subjetivo, uso y administración del tiempo WAIT-2022

A continuación, conteste por favor las siguientes preguntas.

Si ya aparecen datos en las casillas o quiere volver a empezar la encuesta use el siguiente botón: [Borrar los datos y volver a empezar]

Sección 1: Datos personales:

Edad: _______

¿Su estado civil es soltero(a)? ( ) Sí ( ) No

¿Es estudiante? ( ) No soy estudiante ( ) Estudiante, tiempo parcial

( ) Estudiante, dedicación exclusiva

Sección 2: Datos sobre ambientes doméstico

Número de cuartos usados para dormir en su vivienda: _______

Número de personas que viven normalmente en su vivienda: ________

Sección 3: Datos sobre uso de tiempo y bienestar, de semana completa:

Para las preguntas de la sección 3, por favor escriba el tiempo dedicado a cada actividad durante la semana pasada, por separado para entresemana (lunes a viernes) y fin de semana (sábado y domingo).

- De lunes a viernes, y sábados y domingos (horas y/o minutos)
    - Dormir incluyendo siestas.
    - Comer sus alimentos diarios (desayuno, comida, cena, etc.):
    - Aseo o arreglo personal como bañarse, ir al baño, lavarse los dientes, etcétera:
    - Limpiar o recoger el interior de su vivienda (ordenar objetos, tender camas, barrer, trapear, sacudir, lavar la cocina, el baño, entre otros):

  - Entretenerse usando e-mail, chats, redes sociales (WhatsApp, Facebook, Twitter, entre otros) sin hacer otra actividad al mismo tiempo:
  - Realizar actividades artísticas o culturales (tocar un instrumento musical, pintar o realizar artes plásticas, gráficas, literarias o escénicas; incluye danza):
  - Asistir a eventos cívicos o políticos (reuniones, mítines, desfiles, marchas, celebraciones, etc.):
  - Asistir a reuniones sociales o fiestas; o atender visitas de familiares, amigos o conocidos (visitar o recibir a alguien; ir al antro, bar, hablar por teléfono, escribir correspondencia):
  - Hacer voluntariado o participar en organizaciones como asociaciones civiles, Cruz Roja, asilos, casa hogar, DIF, hospitales, iglesias, Alcohólicos Anónimos, partidos políticos, etcétera:
  - Asistir o participar en actividades o celebraciones religiosas:
  - Dedicar un tiempo especial (sin hacer otra actividad) a los integrantes de su hogar para platicar de las actividades diarias, consolar o aconsejar:
    - Para todas las actividades se da la opción de marcar una casilla si se esta de acuerdo con la afirmación:
      - Mejora mi capacidad o habilidad de interactuar con mis ambientes, y cambiarlo si hace falta.

Sección 4: Datos sobre uso de tiempo y bienestar, entre semana:

Para las preguntas de la sección 4, por favor escriba el tiempo dedicado a cada actividad durante la semana pasada, pero únicamente entresemana (lunes a viernes)

- De lunes a viernes (horas y/o minutos)
  - Trabajar:
  - Asistir a clases, tomar cursos o estudiar (incluya sistema abierto o a distancia, diplomados, etcétera):
  - Hacer tareas, prácticas escolares o alguna otra actividad de estudio:
  - Trasladarse de ida y vuelta a escuela o universidad:
  - Cocinar, preparar o calentar alimentos o bebidas:
  - Servir la comida, recoger, lavar, secar o acomodar los trastes:

- Separar, doblar, acomodar o guardar la ropa:
- Cerrar puertas y ventanas, poner candados u otras medidas para proteger sus bienes y su vivienda (guardó el auto, encendió la alarma):
- Hacer deporte o ejercicio físico en su tiempo libre (fútbol, basquetbol, natación, box, karate, caminar, correr, patinar, andar en bicicleta, yoga, zumba):
- Ver películas, novelas, series, programas, videos o documentales en televisión, tablet, celular o computadora (sin hacer otra actividad al mismo tiempo y sin ser por trabajo o estudio):
- Leer algún libro, revista, periódico o artículo mediante algún dispositivo digital o impreso (excluya si es por trabajo o estudio):
    - Para todas las actividades se da la opción de marcar una casilla si se está de acuerdo con la afirmación:
        - Mejora mi capacidad o habilidad de interactuar con mis ambientes, y cambiarlo si hace falta.

Sección 5: Bienestar subjetivo

- ¿Cómo se siente con el tiempo que le dedico la semana pasada a...
    - ...sus clases, cursos o estudios?
    - ...su trabajo remunerado o actividad económica?
    - ...las actividades domésticas que hizo en su hogar?
    - ...hacer lo que realmente le gusta?
    - ...cuidar y apoyar a las personas de su hogar?
    - ...convivir con familiares y amigos?
    - ...los traslados a su trabajo o escuela?
    - ...hacer trámites, pagos o cobrar algún programa social que recibe o recibió?
        - Opciones de respuesta
            - Quisiera dedicarle menos tiempo.
            - Quisiera dedicarle más tiempo.
            - Está bien el tiempo que le dedique.
            - No aplica.

- ¿Como se siente con...
    - ...con su vida en general?
    - ...con su vida familiar?
    - ...con su vida afectiva (el cariño que da y recibe)?
    - ...con su vida social?
    - ...con su situación económica?
    - ...con su vivienda?
        - Opciones de respuesta:
            - Nada satisfecho.
            - Poco satisfecho.
            - Mas o menos satisfecho.
            - Satisfecho.
            - Muy satisfecho.

- En general...
    - ¿Qué tan feliz diría que es usted?
        - Opciones de respuesta:
            - Nada.
            - Poco feliz.
            - Mas o menos feliz.
            - Feliz.
            - Muy feliz.

Annex 2: English translation of the survey

<u>Survey on subjective wellbeing, time use and management WAIT-2022</u>

Please answer the following questions.

If data already appears in the boxes or you want to start the survey again, use the following button: [Clear data and start again]

Section 1: Personal data:

Age: ________

Is your marital status single? ( ) Yes.

Are you a student? ( ) I am not a student ( ) Student, part time

( ) Student, exclusive dedication

Section 2: Data on domestic environments

Number of rooms used for sleeping in your home: ________

Number of people who normally live in your home: _________

Section 3: Time Use and Wellbeing Data, Full Week:

For the questions in section 3, please write the time spent on each activity during the past week, separately for weekdays (Monday to Friday) and weekends (Saturday and Sunday).

- From Monday to Friday, and Saturdays and Sundays (hours and/or minutes)
    - Sleep including naps.
    - Eat your daily meals (breakfast, lunch, dinner, etc.):
    - Cleanliness or personal grooming such as bathing/showering, going to the bathroom, brushing teeth, etc.:
    - Clean or pick up the inside of your home (organize objects, make beds, sweep, mop, dust, wash the kitchen, bathroom, among others):
    - Entertain yourself using email, chats, social networks (WhatsApp, Facebook, Twitter, among others) without doing another activity at the same time:

- Carry out artistic or cultural activities (playing a musical instrument, painting or performing plastic, graphic, literary or performing arts; includes dance):
- Attend civic or political events (meetings, rallies, parades, marches, celebrations, etc.):
- Attend social gatherings or parties; or attend visits from family, friends or acquaintances (visit or receive someone; go to the club, bar, talk on the phone, write correspondence):
- Volunteer or participate in organizations such as civil associations, Red Cross, nursing homes, home, DIF, hospitals, churches, Alcoholics Anonymous, political parties, etc.:
- Attend or participate in religious activities or celebrations:
- Dedicate special time (without doing other activities) to the members of your household to talk about daily activities, console or advise:
  - For all activities, respondents have the option to check a box if they agree with the statement:
    - It improves my capacity or ability to interact with my environments and change it if necessary.

Section 4: Data on time use and well-being, weekdays:

For the questions in section 4, please write the time spent on each activity during the past week, but only on weekdays (Monday to Friday)

- From Monday to Friday (hours and/or minutes)
  - Work:
  - Attend classes, take courses or study (includes open or distance system, diplomas, etc.):
  - Do homework, school practices or some other study activity:
  - Traveling to and from school or university:
  - Cooking, preparing, or heating food or drinks:
  - Serving food, picking up, washing, drying, or arranging dishes:
  - Separate, fold, arrange or store clothes:

- Close doors and windows, put padlocks or other measures to protect your property and your home (store the car, turn on the alarm):
- Do sports or physical exercise in your free time (soccer, basketball, swimming, boxing, karate, walking, running, skating, cycling, yoga, Zumba):
- Watch movies, novels, series, programs, videos or documentaries on television, tablet, cell phone or computer (without doing another activity at the same time and without being for work or study):
- Read a book, magazine, newspaper, or article using a digital or printed device (exclude if it is for work or study):
  - For all activities, respondents have the option to check a box if they agree with the statement:
    - It improves my capacity or ability to interact with my environments and change it if necessary.

Section 5: Subjective well-being

- How do you feel about the time you dedicated last week to ...
  - ...your classes, courses, or studies?
  - ...your paid work or economic activity?
  - ...the domestic activities you did at home?
  - ...do what you really like?
  - ...care for and support the people in your home?
  - ...live with family and friends?
  - ...the commutes to your work or school?
  - ...do paperwork, payments or collect any social program that you receive or received?
    - Response options
      - I would like to spend less time on it.
      - I would like to spend more time to it.
      - Is fine.
      - Does not apply.

- How do you feel with...
    - ...with your life in general?
    - ...with your family life?
    - ...with your emotional life (the affection you give and receive)?
    - ...with your social life?
    - ...with your financial situation?
    - ...with your home?
        - Answer options:
            - Not at all satisfied.
            - Little satisfied.
            - More or less satisfied.
            - Satisfied.
            - Very satisfied.

- In general...
    - How happy would you say you are?
        - Answer options:
            - Nothing.
            - Little happy.
            - More or less happy.
            - Happy.
            - Very happy.

Annex 3: Original text in Spanish of the E-mail sent to students

¡Hola!

Estas recibiendo este correo porque contestaste la Encuesta de Bienestar Subjetivo, uso y Administración del Tiempo WAIT-2022. El sistema de inteligencia artificial ha procesado tus datos en conjunto con el de tus compañeros, y tiene algunas recomendaciones que puedes considerar teniendo en mente tu situación de estudiante universitario.

A continuación, te presentamos listas de actividades que el sistema inteligente recomienda incrementar o disminuir para una mayor satisfacción de uso del tiempo en actividades académicas y en percepción de felicidad en general. Es decir, la predicción del sistema es que, si implementas todos o algunos de estos cambios, en algunas semanas podrás decir que estas más satisfecho con el tiempo que dedicas a actividades académicas.

[Recommendations are inserted here]

Toma en cuenta las relaciones de los patrones de uso de tiempo con la satisfacción y la felicidad son complejas y no necesariamente causales. Actividades que a primera vista no tienen nada que ver con actividades académicas, pueden aumentar la satisfacción en relación con dichas actividades académicas e incluso en general. Por otro lado, no siempre es posible hacer algunos de los cambios recomendados, si este es tu caso, es importante que pienses que tipo de situación no te permite hacer un cambio que es potencialmente beneficioso y conscientemente tomar una decisión al respecto, incluso si esta decisión es no hacer ningún cambio por el momento.

Como una guía general de actividades que se ha demostrado pueden ser beneficiosas para el bienestar personal puedes consultar la siguiente table basada en el modelo de bienestar Wellbeing Five.

[**Table 1** is inserted here]

¡Saludos!,

Equipo de Investigación de Algoritmos inteligentes para administración y uso del tiempo centrados en la preservación y aumento del bienestar humano

------

La información y sus documentos adjuntos son susceptibles de estar clasificados en términos de la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO). Si usted ha recibido este mensaje por error, por favor notifique al remitente respondiendo al correo, y elimine el mensaje original, así como sus documentos adjuntos como una medida de seguridad de carácter administrativo.

This email and its attachments are confidential and may be privileged or protected by law. If you have received it in error, please notify us immediately, delete the email message, do not copy it, disclose its contents or use it for any purpose.

Annex 4: English translation of the E-mail sent to students

Hello!

You are receiving this email because you answered the WAIT-2022 Subjective Wellbeing, Time use and Time Management Survey. The artificial intelligence system has processed your data together with that of your classmates and has some recommendations that you can consider, keeping in mind your situation as a university student.

Below, we present lists of activities that the intelligent system recommends increasing or decreasing for greater satisfaction in the use of time in academic activities and in the perception of happiness in general. That is, the system's prediction is that, if you implement all or some of these changes, in a few weeks you will be able to say that you are more satisfied with the time you spend on academic activities.

[Recommendations are inserted here]

You should consider that the relationships of time use patterns with satisfaction and happiness are complex and not necessarily causal. Activities that at first glance have nothing to do with academic activities, can increase satisfaction in relation to said academic activities and even in general. On the other hand, it is not always possible to make some of the recommended changes. If this is your case, it is important that you think about what type of situation does not allow you to make a change that is potentially beneficial and consciously decide about it, even if it is decision is not to make any changes at the moment.

As a general guide to activities that have been shown to be beneficial for personal well-being, you can consult the following table based on the Wellbeing Five model of well-being.

[**Table 1** is inserted here]

Greetings!

Research Team for Intelligent Algorithms for time management and use focused on the preservation and increase of human well-being

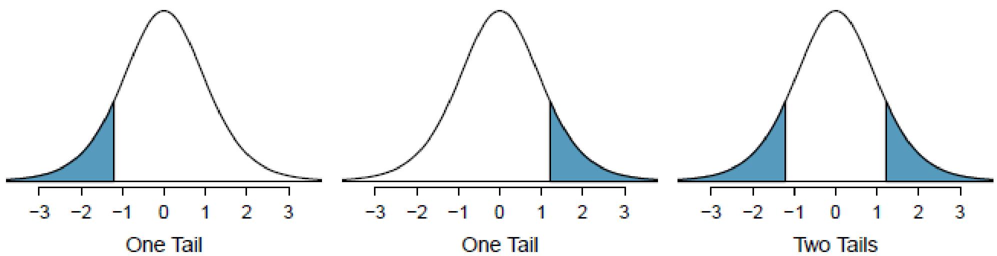
------

The information and its attached documents are likely to be classified in terms of the General Law on Protection of Personal Data Held by Obligated Subjects (LGPDPPSO). If you have received this message in error, please notify the sender by replying to the email, and delete the original message, as well as its attached documents as an administrative security measure.

This email and its attachments are confidential and may be privileged or protected by law. If you have received it in error, please notify us immediately, delete the email message, do not copy it, disclose its contents or use it for any purpose.

Annex 5: T table

This T table was taken from the fourth edition of "OpenIntro Statistics" (Diez et al., 2019).



one tail		0.100	0.050	0.025	0.010	0.005
two tails		0.200	0.100	0.050	0.020	0.010
df	1	3.08	6.31	12.71	31.82	63.66
	2	1.89	2.92	4.30	6.96	9.92
	3	1.64	2.35	3.18	4.54	5.84
	4	1.53	2.13	2.78	3.75	4.60
	5	1.48	2.02	2.57	3.36	4.03
	6	1.44	1.94	2.45	3.14	3.71
	7	1.41	1.89	2.36	3.00	3.50
	8	1.40	1.86	2.31	2.90	3.36
	9	1.38	1.83	2.26	2.82	3.25
	10	1.37	1.81	2.23	2.76	3.17
	11	1.36	1.80	2.20	2.72	3.11
	12	1.36	1.78	2.18	2.68	3.05
	13	1.35	1.77	2.16	2.65	3.01
	14	1.35	1.76	2.14	2.62	2.98
	15	1.34	1.75	2.13	2.60	2.95
	16	1.34	1.75	2.12	2.58	2.92
	17	1.33	1.74	2.11	2.57	2.90
	18	1.33	1.73	2.10	2.55	2.88
	19	1.33	1.73	2.09	2.54	2.86
	20	1.33	1.72	2.09	2.53	2.85
	21	1.32	1.72	2.08	2.52	2.83
	22	1.32	1.72	2.07	2.51	2.82
	23	1.32	1.71	2.07	2.50	2.81
	24	1.32	1.71	2.06	2.49	2.80
	25	1.32	1.71	2.06	2.49	2.79
	26	1.31	1.71	2.06	2.48	2.78
	27	1.31	1.70	2.05	2.47	2.77
	28	1.31	1.70	2.05	2.47	2.76
	29	1.31	1.70	2.05	2.46	2.76
	30	1.31	1.70	2.04	2.46	2.75