



**UNIVERSIDAD AUTONOMA
DE AGUASCALIENTES**

Centro de Ciencias Básicas

Departamento de Sistemas de Información

Tesis

**Identificación de patrones vinculados al desempeño estudiantil en el nivel básico
demostrado en la prueba EXCALE.**

Presenta

Ing. Juan Alberto Moreno Sánchez

Para obtener el grado de maestro en informática y tecnologías computacionales

Comité tutorial

Tutor: Dra. María Dolores Torres Soto

Co-Tutora: Dra. Aurora Torres Soto

Asesor: Dr. Cesar Eduardo Velázquez Amador

Aguascalientes, Ags., mayo 2020

Fecha de dictaminación dd/mm/aa: 23/06/20

NOMBRE: JUAN ALBERTO MORENO SÁNCHEZ **ID** 166859

PROGRAMA: MAESTRÍA EN INFORMÁTICA Y TECNOLOGÍAS COMPUTACIONALES **LGAC (del posgrado):** La Ingeniería / Enfoque de sistemas para el mejoramiento de los procesos organizacionales usando SI/TI.

TIPO DE TRABAJO: () Tesis () Trabajo práctico

TITULO: Identificación de patrones vinculados al desempeño estudiantil en el nivel básico demostrado en la prueba EXCALE

IMPACTO SOCIAL (señalar el impacto logrado): IMPACTO EN EL SISTEMA EDUCATIVO NACIONAL DE NIVEL BÁSICO

INDICAR SI/NO 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, tecnológica 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
- SI Generó transferencia del conocimiento o tecnológica
- SI Cumpe 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 tutorial, en caso de los posgrados profesionales si tiene solo tutor podrá 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
- N/A 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 artículos 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 ADSCRIPCIÓN:

DR. JOSÉ MANUEL MORA TAVAREZ

NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO:

MITC. JORGE EDUARDO MACÍAS LUEVANO

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

Revisó:

NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:

DRA. en C. BGO. HAYDEE MARTÍNEZ RUIVALCABA

Autorizó:

NOMBRE Y FIRMA DEL DECANO:

M. en C. JORGE MARTÍN ALFÉREZ CHÁVEZ

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

En cumplimiento con el Art. 205C 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. 205F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.



M. en C. Jorge Martín Alférez Chávez
DECANO DEL CENTRO DE CIENCIAS BASICAS

PRESENTE

Por medio del presente como **TUTOR** designado del estudiante **JUAN ALBERTO MORENO SÁNCHEZ** con ID 166859 quien realizó el trabajo de tesis titulado: **IDENTIFICACIÓN DE PATRONES VINCULADOS AL DESEMPEÑO ESTUDIANTIL EN EL NIVEL BÁSICO DEMOSTRADO EN LA PRUEBA EXCALE**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que *él* pueda proceder a imprimirlo, así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE

"Se Lumen Proferre"

Aguascalientes, Ags., a 15 de junio de 2020.

Dra. María Dolores Torres Soto
Tutor de trabajo de tesis

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



UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES

CARTA DE VOTO APROBATORIO
INDIVIDUAL

M. en C. Jorge Martín Alférez Chávez
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como **CO-TUTOR** designado del estudiante **JUAN ALBERTO MORENO SÁNCHEZ** con ID 166859 quien realizó el trabajo de tesis titulado: **IDENTIFICACIÓN DE PATRONES VINCULADOS AL DESEMPEÑO ESTUDIANTIL EN EL NIVEL BÁSICO DEMOSTRADO EN LA PRUEBA EXCALE**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que *él* pueda proceder a imprimirlo, así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE
"Se Lumen Proferre"

Aguascalientes, Ags., a 15 de junio de 2020.

Dra. Aurora Torres Soto
Co-tutor de trabajo de tesis

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

M. en C. Jorge Martín Alférez Chávez
DECANO DEL CENTRO DE CIENCIAS BASICAS

PRESENTE

Por medio del presente como **ASESOR** designado del estudiante **JUAN ALBERTO MORENO SÁNCHEZ** con ID 166859 quien realizó el trabajo de tesis titulado: **IDENTIFICACIÓN DE PATRONES VINCULADOS AL DESEMPEÑO ESTUDIANTIL EN EL NIVEL BÁSICO DEMOSTRADO EN LA PRUEBA EXCALE**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que él pueda proceder a imprimirlo, así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE
"Se Lumen Proferre"
Aguascalientes, Ags., a 15 de junio de 2020.



Dr. César Eduardo Velázquez Amador
Asesor de trabajo de tesis

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

AGRADECIMIENTOS

A la Universidad Autónoma de Aguascalientes por su apoyo a lo largo de este proyecto. Al CONACYT que, sin su apoyo, lograr esta meta hubiera sido más complicado.

A mis tutores la Dra. Dolores, la Dra. Aurora y el Dr. Cesar por el apoyo, dedicación, pero sobre todo paciencia; por todo lo que he aprendido de cada uno de ustedes; de la universidad Autónoma son un ejemplo de dedicación y atención y como personas un ejemplo de amabilidad y respeto. Muchas Gracias.

A mi familia por estar conmigo incondicionalmente, y ser siempre la motivación de todas mis actividades.

A mis viejos y nuevos amigos que desde hace muchos años son testigos y un gran soporte para cumplir aquellas cosas que me propongo.

A cada una de las personas que, de alguna u otra manera, se involucraron en la realización de este proyecto, así como aquellos que me alentaron y creyeron en mí en momentos en los que, incluso, yo mismo desistía. Muchas gracias.

DEDICATORIAS

“A la única persona que realmente merece una dedicatoria,
mi mamá”



Índice general

Índice de tablas	3
Índice de figuras.....	4
Índice de algoritmos.....	4
RESUMEN	5
ABSTRACT.....	7
1. INTRODUCCIÓN.....	8
1.1. Problema de investigación	9
1.2. Justificación.....	10
1.3. Objetivos del proyecto	11
1.3.1. Objetivo general.....	11
1.3.2. Objetivos específicos	11
1.4. Hipótesis.....	12
2. MARCO TEÓRICO	13
2.1. Antecedentes	13
2.1.1. Estudios sobre los factores que influyen en el desempeño escolar.....	13
2.2. El rol del INEE en el estudio de factores asociados en México.....	15
2.3. Importancia de los estudios sobre factores asociados al aprendizaje.....	16
2.3.1. Importancia de la educación en México y América Latina	17
2.3.2. Relación entre calidad educativa y crecimiento económico	20
2.3.3. El INEE como herramienta en la mejora educativa.....	22
2.3.4. La prueba EXCALE en la educación en México.....	22
2.4. Reconocimiento de Patrones.....	26
2.4.1. Objetivos de los enfoques de reconocimiento de patrones	27
2.4.2. Problemas que ataca el Reconocimiento de Patrones	29
2.5. Testores y Testores Típicos.....	30
3. METODOLOGÍA.....	32
3.1. Actividades Principales.....	32
3.2. Recolección de información.....	34
3.3. Preprocesamiento	35

3.3.1.	Transformación de variables de trabajo en la base de datos de la evaluación de contexto.....	36
3.3.2.	Transformación de las variables de trabajo de la base de datos de resultados en la prueba de matemáticas.....	46
3.3.3.	Datos de entrada para el algoritmo	50
3.3.4.	Equipos y herramientas.....	51
3.4.	Implementación del algoritmo	52
3.4.1.	Elementos del algoritmo desarrollado	53
3.5.	Construcción de la herramienta para validación	61
4.	RESULTADOS	64
4.1.	Identificación de factores para la matriz de entrada.....	64
4.1.1.	Análisis de resultados para la matriz de entrada.....	65
4.2.	Identificación de factores para la matriz de entrada incluyendo el campo entidad	65
4.2.1.	Análisis de resultados para la matriz de entrada incluyendo la variable entidad	67
4.3.	Identificación de factores que determinan el desempeño de un alumno en la prueba EXCALE en el estado de Aguascalientes	67
4.3.1.	Identificación de testores típicos en el estado de Aguascalientes.....	67
4.3.2.	Análisis de resultados para el estado de Aguascalientes	68
4.4.	Identificación de factores que determinan el desempeño de un alumno en la prueba EXCALE en el estado de Jalisco	71
4.4.1.	Identificación de testores típicos en el estado de Jalisco	71
4.4.2.	Análisis de resultados para el estado de Jalisco.....	72
4.5.	Identificación de factores que determinan el desempeño de un alumno en la prueba EXCALE en el estado de Veracruz.....	74
4.5.1.	Identificación de testores típicos en el estado de Veracruz	74
4.5.2.	Análisis de resultados para el estado de Veracruz	75
5.	CONCLUSIONES.....	77
5.1.	Objetivos cubiertos.....	78
5.2.	Contribuciones	80
5.3.	Trabajo futuro.....	82
6.	BIBLIOGRAFÍA	83

Índice de tablas

Tabla 1 Niveles, Grados y asignaturas a evaluar (INEE, 2005)	23
Tabla 2 Aplicaciones de los Excale por año (Zamudio Mesa, Díaz Argüero, & Lepe García, 2012)	24
Tabla 3 Constructo Edad Normativa (C1)	36
Tabla 4 Constructo Vivienda (C2).....	37
Tabla 5 Constructo Situación Económica (C3)	38
Tabla 6 Constructo Contexto Padres (C4)	40
Tabla 7 Constructo Futuro Intelectual (C5).....	40
Tabla 8 Constructo Entorno Casa Escuela (C6)	41
Tabla 9 Constructo Lengua Indígena (C7)	41
Tabla 10 Constructo "condiciones de estudio" (C8).....	42
Tabla 11 Constructo Actividades Extracurriculares (C9).....	44
Tabla 12 Constructo Antecedentes Académicos (C10)	45
Tabla 13 Ejemplo de valores obtenidos por constructo.....	46
Tabla 14 Logros académicos evaluados en la prueba Excale	47
Tabla 15 Valores posibles en los diferentes logros académicos	49
Tabla 16 Clasificación de alumnos en base a su desempeño.....	50
Tabla 17 Ejemplo de datos de entrada para el algoritmo de testores.....	51
Tabla 18 Datos de entrada para la ejecución del algoritmo.	64
Tabla 19 Testores obtenidos en la ejecución del algoritmo.....	64
Tabla 20 Datos de entrada para la ejecución del algoritmo (incluyendo al campo entidad)	66
Tabla 21 Testores obtenidos en la ejecución del algoritmo (incluyendo al campo entidad).	66
Tabla 22 Testores obtenidos en la ejecución correspondiente al estado de Aguascalientes.	68
Tabla 23 Porcentaje de clasificación en KNIME en el estado de Aguascalientes.	69
Tabla 24 Testores obtenidos en la ejecución correspondiente al estado de Jalisco	71
Tabla 25 Porcentaje de clasificación en KNIME en el estado de Jalisco.	72
Tabla 26 Testores obtenidos en la ejecución correspondiente al estado de Veracruz.	74
Tabla 27 Porcentaje de clasificación en KNIME en el estado de Veracruz.	75

Índice de figuras

Figura 1 Nivel de salario con respecto a su formación académica (Hanushek & Ludger 2007)	20
Figura 2 Salario por hora de los trabajadores adultos y jóvenes.	21
Figura 3 Enfoques del Reconocimiento de Patrones / Ubicación de la Selección de Características (Acosta, 2018)	27
Figura 4 Conjuntos de Características (Acosta, 2018)	31
Figura 5 Metodología utilizada.....	32
Figura 6 Etapas del algoritmo de Testores.....	52
Figura 7 Herramienta de validación creada en KNIME.	61
Figura 8 Árbol bayesiano del estado de Aguascalientes.....	70
Figura 9 Árbol bayesiano para el estado de Jalisco.	73
Figura 10 Árbol bayesiano para el estado de Veracruz	76

Índice de algoritmos

Algoritmo 1 Eliminación de contradicciones	55
Algoritmo 2 Cálculo de matriz de diferencias	56
Algoritmo 3 Cálculo de matriz de diferencias simples.....	57
Algoritmo 4 Cálculo de matriz Básica.....	57
Algoritmo 5 Cálculo de Testores	58
Algoritmo 6 Cálculo de Testores Típicos	59

RESUMEN

Debido a la evolución que ha sufrido la información digital y el análisis de esta información, hoy en día existen un sin número de bases de datos de gran tamaño y en ramos diferentes de la investigación. Un caso específico es la educación, que, a través de sus procesos de evaluación, provee grandes cantidades de información digital. Por esta razón, la interacción de la educación con las ciencias computacionales resulta en una novedosa herramienta que permite vislumbrar las causas del comportamiento, desempeño y otras situaciones que ocurren en el sistema educativo.

El siguiente documento describe la aplicación de la lógica combinatoria, también llamada teoría de testores, encontrando aquellos conjuntos de características que inciden de forma determinante en el desempeño educativo de los alumnos en la prueba EXCALE. En concreto, se analizó la base de datos de los resultados de matemáticas en la prueba EXCALE_03_2014 clasificando el desempeño de los alumnos en alto, medio y bajo en base a los resultados obtenidos en esta prueba. Para este análisis se desarrolló y utilizó una metodología que involucrara la teoría de testores que permite identificar conjuntos de testores típicos, los cuales representan la información mínima necesaria a partir de sus características para distinguir objetos en sus respectivas clases.

La identificación de testores típicos es un problema exponencial respecto al número de características involucradas. Por esta razón, se generaron agrupaciones de características llamadas constructos, que minimizaran el procesamiento de la información y a su vez, garantizaran la integridad de esta información.

La implementación de esta metodología se realizó para diferentes estados de la república mexicana, utilizando un porcentaje de la información para la identificación de características determinísticas y el resto de la información para evaluar la eficacia de los testores típicos en la clasificación.

Al final de esta investigación se contó con una metodología capaz de identificar las características que definen el desempeño educativo de los alumnos en la prueba EXCALE, así como dichas características correspondientes a diferentes grupos de información

(Estados), además se construyeron árboles bayesianos que definen el comportamiento que siguen los diferentes desempeños educativos en base a sus características.



ABSTRACT

Due to the evolution that digital information has undergone and the analysis of this information, today there are countless large databases in different branches of research. A specific case is education, which, through its evaluation processes, provides large amounts of digital information. For this reason, the interaction of education with computer science results in a novel tool that allows us to glimpse the causes of behavior, performance and other situations that occur in the educational system.

The following document describes the application of combinatorial logic, also called testors theory, finding those sets of characteristics that have a decisive impact on the educational performance of students on the EXCALE test. Specifically, the database of the results of mathematics in the EXCALE_03_2014 test was analyzed, classifying the performance of the students in high, medium and low based on the results obtained in this test. For this analysis, a methodology involving testors theory was developed and used to identify typical testors sets, which represent the minimum information necessary based on their characteristics to distinguish objects in their respective classes.

The identification of typical testors is an exponential problem with respect to the number of characteristics involved. For this reason, groups of characteristics called constructs were generated, which minimized the processing of information and, in turn, guaranteed the integrity of this information.

The implementation of this methodology was carried out for different states of the Mexican Republic, using a percentage of the information to identify deterministic characteristics and the rest of the information to evaluate the effectiveness of typical testors in the classification.

At the end of this research, there was a methodology capable of identifying the characteristics that define the educational performance of the students in the EXCALE test, as well as those characteristics corresponding to different groups of information (States). Bayesian trees were also defined that define the behavior that different educational performances follow based on their characteristics.

1. INTRODUCCIÓN

El presente documento de tesis aplica el enfoque lógico-combinatorio del reconocimiento de patrones a los resultados de matemáticas en la prueba EXCALE, por medio de la selección de constructos y de la aplicación de la teoría de testores haciendo uso del algoritmo exhaustivo. El uso de este enfoque permitió analizar un porcentaje de la base de datos y determinar qué características definen el desempeño educativo de los alumnos, mientras que el uso de un modelo de clasificación desarrollado en KNIME ayudó a evaluar y validar la eficacia de los testores típicos mediante la creación de árboles de decisión bayesianos que explican el comportamiento de cada clase de desempeño en base a sus características.

Este trabajo se compone de seis apartados además de la introducción, de los cuales el segundo apartado aborda el marco teórico con los conceptos básicos que fueron necesarios para la realización de este documento. En el tercer apartado se describe la metodología que se siguió y las características de su aplicación donde se incluyen una explicación de la implementación del algoritmo de testores típicos, así como las diferentes fases y componentes de este algoritmo, por otro lado, en este mismo apartado se muestra la composición y el desarrollo en KNIME de la herramienta de evaluación y validación de los testores típicos. Siguiendo con el cuarto apartado, se exponen los resultados obtenidos, los testores típicos, su representación característica, así como la medición de su eficacia y las reglas de clasificación de cada modelo de datos. Finalmente, las conclusiones se abordan en el quinto y último apartado, donde se explica que representan los diferentes resultados en relación con el desempeño educativo de la prueba EXCALE, así como aquellas contribuciones que aporta esta investigación en el área educativa y el área de inteligencia artificial, en este apartado se explican también posibles trabajos futuros que se pudieran realizar a partir de esta investigación.

1.1. Problema de investigación

La educación, como factor de desarrollo de un país, se ha usado para mejorar la calidad de vida de sus habitantes, asignando recursos en investigación al servicio de su mejora y de su desarrollo. Sin embargo, en la última década, los avances en educación en varios países, en el nivel básico, se han visto mermados o no han tenido un progreso significativo.

Lo anterior, ha generado consecuencias negativas para el desarrollo de estas naciones. Debido a que la educación es uno de los pilares fundamentales en el desarrollo social de un país, su carencia o su mala calidad afectan directamente en muchos otros sectores nacionales como lo son la seguridad, la economía, la productividad, su comportamiento social y sin dejar atrás a su política interior.

En el caso de México, se han realizado grandes esfuerzos para lograr una mejora educativa mediante la implementación de políticas públicas destinadas a la mejora del sector educativo, programas para fortalecer los sectores vulnerables, la creación de institutos dedicados a la evaluación y al análisis de la educación. No obstante, lo anterior, los índices educativos no han tenido el avance o la mejora esperados. Podemos decir que, en México, aunque seamos conscientes del nivel educativo actual, las causas de esto, no son completamente claras, es decir, que no conocemos claramente los factores que afectan su desempeño.

Lo anterior, dificulta el proponer soluciones e implementar estrategias realmente funcionales. Según el Instituto Nacional Para la Evaluación de la Educación, existen 25, 666, 451 estudiantes de educación básica en México (INEE, 2012) de los cuales al entrar a la primaria, el 10% no sabe que se escribe de izquierda a derecha y al salir de primaria, el problema se agrava, “cerca de 15% de los alumnos no puede localizar información en documentos como directorios telefónicos o planos de una ciudad, un 12%, no puede resolver operaciones de multiplicación y división con números enteros; 27% no reconoce que los alimentos en mal estado, las bebidas alcohólicas y el tabaco son factores que alteran el funcionamiento del cuerpo; es preciso mejorar la educación en México y con ello las condiciones de vida de sus habitantes. Es por esto que resulta relevante valerse de los avances

en ciencia y tecnología para realizar estudios que ayuden a identificar los factores que afectan directamente el desempeño educativo de los alumnos en educación básica y partir de estos resultados, cimbrar las bases para una mejora notoria y consistente con nuestra realidad.

1.2. Justificación

Considerando que la educación afecta directa e indirectamente en el desarrollo de la gran mayoría de los sectores de un país, es necesario identificar los factores que determinan el desempeño educativo de los alumnos. Con este insumo, se podrá crear un marco que permita implementar estrategias para corregir factores que caracterizan el fracaso educativo, o en su caso, replicar factores que influyen en el éxito educativo.

Es aquí donde el enfoque lógico combinatorio conocido como teoría de testores, nos puede apoyar para determinar factores de éxito educativo a nivel básico.

Si logramos identificar los elementos que caracterizan el éxito educativo de los estudiantes de nivel básico en México, podremos proponer la réplica de estos factores en diferentes sectores educativos del país, por otro lado, si determinamos aquellas características que determinan un bajo desempeño educativo, es posible generar estrategias que permitan enmendar y/o prevenir esta situación.

Un testor es un conjunto de características que definen la descripción de objetos que pertenecen a diferentes clases. Los testores ayudan a encontrar una solución en la identificación de patrones de objetos pertenecientes a diferentes clases.

Los testores típicos son fundamentales en la selección de variables debido a que nos proporciona la capacidad de saber si existen grupos de variables que puedan determinar la clase a la que pertenece un objeto.

Sabemos que existen diversos análisis estadísticos que proveen herramientas para la identificación de patrones, sin embargo, estos análisis tienen ciertas limitaciones como lo es contar con conjuntos de datos que tengan distribuciones de probabilidad conocidas y comprobadas como lo es la distribución normal.

Ya que en el mundo real existen una gran cantidad de grupos de datos que no cumplen esta propiedad de pertenecer a una distribución de probabilidad específica, se vuelve indispensable utilizar otro tipo de herramientas para llevar a cabo la identificación de patrones, como en este caso el análisis lógico combinatorio.

1.3. Objetivos del proyecto

1.3.1. Objetivo general

diseño e implementación de un marco de trabajo con base en teoría de testores y redes bayesianas para identificar los factores que impactan en el desempeño educativo de los sustentantes de la prueba EXCALE, en su edición 2014, específicamente en la modalidad de matemáticas, para crear clasificadores que ayuden a definir y predecir el comportamiento de los resultados de la evaluación EXCALE.

1.3.2. Objetivos específicos

- Definir una metodología de análisis de datos acorde a la información disponible.
- Diseñar, desarrollar y probar una herramienta que permita encontrar el 100% de los testores típicos asociados a la base de datos de EXCALE.
- Analizar los resultados arrojados por la herramienta desarrollada e identificar las características (patrones) que definan a las diferentes clases de datos definidas.
- Diseñar una herramienta de clasificación que permita definir y predecir el desempeño de los sustentantes en la prueba EXCALE.
- Evaluar e interpretar los resultados obtenidos de la herramienta de clasificación.

1.4. Hipótesis

Es factible construir e implementar un marco de trabajo que utilice testores típicos y redes bayesianos que provea un mecanismo confiable para obtener conjuntos de características que distinguen a los estudiantes de alto, medio y bajo desempeño académico en la prueba Excale de Matemáticas en su edición 03_2014.



2. MARCO TEÓRICO

En este apartado se explican antecedentes de estudios sobre características que influyen en el desempeño escolar, así mismo, se explica la función que tiene el INNE dentro de la educación en México, y también la importancia que tienen los estudios que tratan de explicar los factores que están directamente asociados al aprendizaje académico. Finalmente se describen cada uno de los conceptos fundamentales que respaldaron la realización de este documento de tesis como los son: los testores y testores típicos, la selección de características y el reconocimiento de patrones, así como sus objetivos y los problemas que ataca.

2.1. Antecedentes

2.1.1. Estudios sobre los factores que influyen en el desempeño escolar.

A finales de la década de los 60, el informe Coleman (Hopkins & Coleman, 1966) llamó profundamente la atención sobre la importancia de los factores estructurales en el desempeño académico de los estudiantes. El origen social, la clase, la zona de residencia y ciertos aspectos culturales de las familias y de la localidad más inmediata a los mismos, fueron los aspectos que mejor explicaban las diferencias en el desempeño académico.

Inmediatamente después del informe de Coleman, varios trabajos contribuyeron a confirmar el enorme peso de los factores estructurales en el rendimiento académico. (Bowles & Gintis, 1985). Ésto generó un enorme pesimismo sobre la capacidad de las escuelas para revertir el efecto de las variables estructurales. El cambio que permitiera mejorar el rendimiento escolar de los sectores sociales menos beneficiados, sólo podría provenir de la transformación de la estructura social.

Las múltiples investigaciones que se realizaron durante esas décadas, fincaron principalmente su atención en las grandes desigualdades educativas existentes entre los diferentes estratos de la población, en cuanto al acceso, la reprobación y la deserción en los distintos niveles educativos, las diferencias en la infraestructura de los establecimientos escolares a los que asistían, o los recursos docentes y pedagógicos que los atendían, constituían las principales variables dependientes en sus análisis.

A partir de estos estudios, se acuñaron diversos términos como: desigualdades educativas, desigualdad de oportunidades educativas, movilidad intergeneracional de las desigualdades educativas, marginación y exclusión educativa, fracaso escolar, etcétera.

En América Latina, diversas investigaciones aportaron evidencias acerca de los efectos generados por los factores y los procesos que, al interior de las escuelas, intervienen en la distribución de las oportunidades educativas. Entre ellas, Muñoz Izquierdo, C. (1996), distingue dos tipos de investigaciones:

1. Los estudios de naturaleza correlacional, basados en la observación de las interacciones entre la distribución de las oportunidades educativas y los factores extrínsecos e intrínsecos al sistema escolar.
2. Los estudios de naturaleza interpretativa, cuyo propósito era reconstruir, a partir de la aplicación de diferentes perspectivas (sistémica, etnográfica, sociológica y psicosocial) las variables y procesos que permitirían explicar las diferencias en la distribución de dichas oportunidades.

En el primer grupo de investigaciones, las principales variables e indicadores relacionados con el logro escolar detectados a través de análisis estadístico multivariante corresponden a los siguientes factores:

Factores extrínsecos al sistema escolar:

- a) Características sociales, económicas y culturales de las familias: ocupación del padre, escolaridad de los padres. (Muñoz Izquierdo & Guzmán, 1971)

Factores intrínsecos al sistema escolar:

- b) Características de las escuelas: disponibilidad de libros de texto, tareas para realizar en casa, tamaño de los grupos, equipamiento escolar, tamaño de la escuela, diseño y operación de los servicios ofrecidos por el sistema escolar, frecuencia de cambios de los profesores durante el año escolar (Muñoz Izquierdo & Guzmán, 1971).
- c) Características de los maestros: formación profesional, experiencia docente, calidad de las interacciones entre los alumnos y el maestro (Muñoz Izquierdo & Guzmán, 1971).

- d) Características de los alumnos: condición socioeconómica, asistencia a preescolar, repetición, sexo, edad, nutrición, peso y acceso a la televisión. (Muñoz Izquierdo & Guzmán, 1971)

2.2. El rol del INEE en el estudio de factores asociados en México

Por la relevancia social y económica que tiene la educación para México, en los últimos años, el INEE ha realizado un gran esfuerzo para conocer los factores que se asocian con el logro educativo de los estudiantes. Dos de sus publicaciones más recientes, con alumnos de sexto de primaria y de tercero de secundaria, donde se analizan los resultados de los exámenes EXCALE (Exámenes de la Calidad y el Logro Educativos) de español y de matemáticas, son:

1. Aprendizaje y desigualdad social en México: el caso de la educación básica (Backhoff Escudero, Bouzas Riaño, Hernández, & García, 2007)
2. Factores escolares y aprendizaje en la educación básica en México (Backhoff Escudero, Bouzas Riaño, Hernández, & García, 2007).

El primero de ellos planteó cinco grandes objetivos:

- a) Establecer la magnitud de las diferencias en el aprendizaje del español y las Matemáticas entre estudiantes que terminan la primaria y la secundaria en México, y estimó qué tanto de estas diferencias pueden atribuirse a desigualdades entre estudiantes, escuelas y entidades federativas.
- b) Conocer el impacto que tiene el nivel sociocultural de los estudiantes sobre estos aprendizajes.
- c) Conocer la fuerza de esta relación en las treinta y dos entidades federativas del país.
- d) Probar o refutar la validez de cuatro hipótesis sobre el logro educativo que se derivan de un análisis detallado del comportamiento del gradiente sociocultural.
- e) Evaluar el éxito del sistema educativo en reducir las inequidades en oportunidades educativas y hacer recomendaciones de política pública que se fundamenten en los hallazgos.

Los resultados de dicho estudio, muestran con toda claridad que las condiciones familiares y sociales de los estudiantes impactan fuertemente en su aprendizaje, calculando que dichas condiciones explican alrededor del 60% de las diferencias en el logro educativo de Español y de Matemáticas; asimismo, confirman para el caso de México, lo ya sabido desde el informe de Coleman y colaboradores; que la distribución del aprendizaje depende del nivel socioeconómico de las familias de los estudiantes, y que la escuela reproduce las inequidades sociales y económicas del país.

El segundo estudio, relacionado con los factores de la escuela, se propuso responder tres preguntas:

1. ¿Qué tan desigual es la educación en México en términos del logro educativo que alcanzan los alumnos en los distintos centros escolares?
2. ¿En qué medida, las diferencias en las puntuaciones promedio de las escuelas pueden explicarse por las variables de proceso escolar?
3. ¿Cuáles variables de la escuela influyen en el aprendizaje y en qué medida lo hacen, una vez que se toman en cuenta las características de los estudiantes?

Los resultados de este trabajo, muestran que la distancia que separa el logro educativo de los estudiantes entre una escuela y otra, puede llegar a ser equivalente a más de cuatro grados escolares; además muestra que las características propias de las escuelas (modalidad educativa, características estructurales, etc.) pueden llegar a explicar una buena proporción de la varianza de los resultados de aprendizaje y que los factores que mayor impacto tienen en el logro educativo, son los relacionados con las características individuales y familiares de los estudiantes, seguidos de la modalidad educativa y los factores de composición de la escuela, e incluso, con menor fuerza, las características estructurales de los centros escolares.

2.3. Importancia de los estudios sobre factores asociados al aprendizaje

Como lo mencionaron algunos autores, por ejemplo: Casanova en 1992, y lo ha sostenido muchas veces el INEE en sus diversos reportes, la evaluación de los centros escolares constituye un elemento indispensable para la mejora de la calidad de la educación. Esta perspectiva, es complementaria a la de diagnosticar los servicios educativos de un país, a través de la valoración de los niveles de logro educativo alcanzados por sus alumnos.

Cuando la evaluación se orienta hacia los centros escolares, se pone el énfasis en aspectos tales como: la infraestructura institucional, procesos educativos, modelos pedagógicos, métodos didácticos, relaciones interpersonales, utilización de recursos educativos, implementación del currículo, sólo por citar algunos de ellos.

En el informe presentado en Washington en 1966 (Hopkins & Coleman, 1966), se comenta que los procesos de evaluación, permiten dar información sobre lo que ocurre al interior de las escuelas y, con ello, comprender cómo lo educativo se puede aprender de una manera más concreta.

Por lo anterior, es necesario destacar la importancia para realizar diversos análisis que contextualicen la educación dentro del difícil mundo del desarrollo de un país. En este sentido, es importante destacar la compleja relación existente entre educación y desarrollo, la que se encuentra acotada por factores que la condicionan. Así, los factores en la educación, por ejemplo: inestabilidad política, pobreza, corrupción y desigualdad económica, pueden anular su eficacia, por lo que hay que tomarlos muy en cuenta. La investigación en el área, ha mostrado que el nivel socioeconómico de los estudiantes es la variable, que se encuentra más asociada a diferentes indicadores de logro educativo (Hopkins & Coleman, 1966).

2.3.1. Importancia de la educación en México y América Latina

Distintos organismos internacionales, en distintas épocas y con distintos instrumentos legales, han reconocido la educación como un derecho innegable para la humanidad, la Declaración Universal de los Derechos Humanos (Asamblea General de las Naciones Unidas, Artículo 26, 1948), el Pacto Internacional de los Derechos Económicos, Sociales y Culturales (COPREDEH, 2011), y la declaración de los Derechos del Niño (Asamblea General de las Naciones Unidas, declaración de los Derechos del Niño, 1959), son algunos de los instrumentos que tienen como objetivo garantizar el derecho a la educación sin discriminación alguna.

Lo anterior, además de resaltar la importancia de este derecho a la educación, también señala la existencia de una preocupación especial por parte de la comunidad internacional respecto de la cantidad de individuos que tienen acceso a la educación y la calidad de esta educación. Fue gracias a esta preocupación internacional, que, en el siglo actual, pese a un evidente estancamiento educativo en América Latina, se observa un aumento significativo en la escolaridad registrada, con respecto al aumento de la densidad poblacional (Simon, 1995).

Si bien es cierto que es importante contar con más individuos con acceso a la educación, también es cierto que resulta igualmente importante, poner atención en la calidad con la que se imparte la misma. La calidad determina no sólo cuanto y qué aprenden los individuos, sino también en qué medida su aprendizaje contribuye a su beneficio personal, y al de la sociedad (UNESCO, 2005).

De acuerdo con la Organización para la Cooperación y el Desarrollo Económico (OCDE), la educación se define como: “La comunicación organizada y sustentada que tiene por objeto producir aprendizaje”. Según el diccionario de la Real Academia Española se puede entender por calidad: “toda propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor”. Por lo tanto, la calidad de la educación resulta ser: toda propiedad inherente a la comunicación organizada y sustentada, cuyo objeto es producir aprendizaje, que permite juzgar su valor.

A diferencia de épocas antiguas, donde la educación era un lujo o algo innecesario para el crecimiento económico nacional, hoy sabemos con exactitud que existe una relación muy estrecha entre la calidad educativa de un país y el desarrollo de éste, además, se ha comprobado que existe una íntima relación entre desarrollo y escolaridad. Los países desarrollados son aquellos que más gastan en educación, por tanto, su población tiene un nivel de escolaridad y de logro académico más alto. A lo largo del tiempo, estos países han comprobado que la educación resulta en un factor clave para el crecimiento económico y el desarrollo a largo plazo, por su contribución a mejorar la calidad de vida del capital humano de su economía (Barro, 2013).

Debido a estas relaciones entre la educación y el desarrollo de un país, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (abreviado internacionalmente como UNESCO) ha desarrollado investigaciones a nivel mundial que ayudan a explicar la relación entre ellos en términos de nivel de logro académico y por otro lado el nivel de desarrollo de los países. La UNESCO, define este estudio en dos grupos, el origen social de los alumnos (referente a ocupaciones e ingresos familiares) y, por otro lado, las características organizacionales de las instituciones que imparten educación (referente a la preparación docente, material educativo e infraestructura con la que cuentan).

Dichos estudios, ejemplifican que uno de los problemas principales que se presentan, es la disparidad en la calidad educativa que ofrecen los diferentes países alrededor del mundo e incluso, la disparidad dentro de los mismos países (Braslavsky & Cosse, 2003).

Para lograr garantizar la equidad educativa dentro de los sistemas escolares, se requieren varias cosas, dentro de las cuáles se encuentra el garantizar la cobertura escolar, es decir, que exista la infraestructura suficiente para cubrir la demanda educativa; una vez cubierta esta demanda se debería continuar con el impulso a la equidad en cuanto a condiciones de aprendizaje se refiere, ésto, con el objetivo de lograr que los estudiantes puedan alcanzar resultados similares sin importar el lugar donde asistan a la escuela.

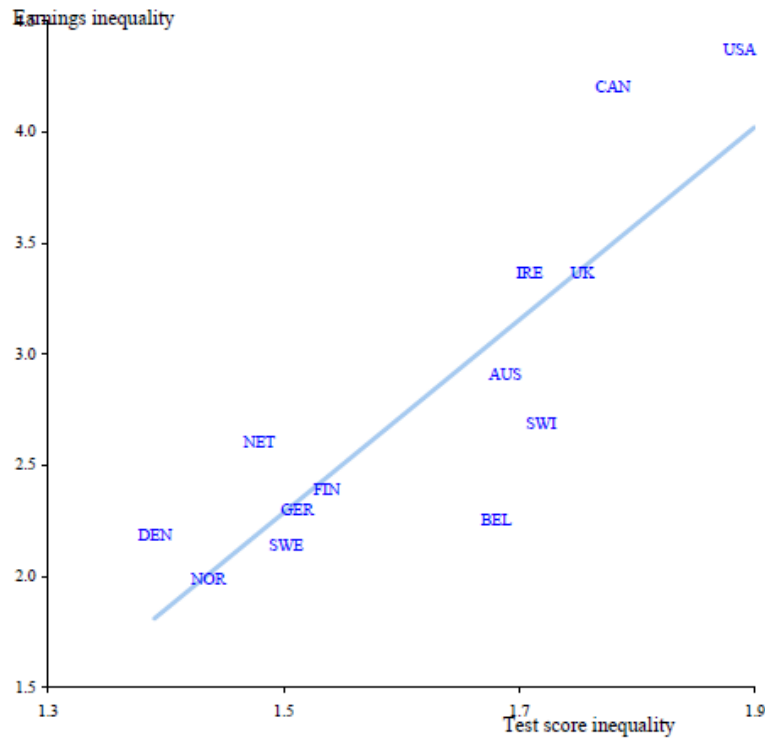
Por lo anterior, se puede afirmar que la educación repercute en la calidad de vida y oportunidades que tendrá el alumno (Castañeda, 16 Mayo 2013) . La calidad educativa cobra especial relevancia en nuestra actual época de globalización, ya que se han incrementado la competencia y la complejidad del mercado laboral, por lo que en la mayoría de los países es cada vez más complicado conseguir un empleo bien remunerado (Lagunes & Maldonado Márquez, 2016).

Es evidente que, en México, en la actualidad, los esfuerzos realizados, aún no han sido suficientes para garantizar el acceso a la educación y que, además, es necesario que los sistemas educativos cuenten con un nivel de calidad educativa que brinde a los alumnos los conocimientos y herramientas necesarias para asegurarles un buen nivel de vida en el futuro.

2.3.2. Relación entre calidad educativa y crecimiento económico

Cuando se tienen alumnos con educación de calidad, éstos, son capaces de adquirir distintas competencias, las cuales les permitirán en un futuro tener un mejor estilo de vida, ya que estas habilidades serán también las que determinarán su nivel de ingresos.

Un estudio realizado por Nickell en el 2014, demostró cómo la distribución de los salarios en distintas partes del mundo, se ve afectada por las diferentes habilidades que poseen los individuos. La Figura 2 muestra los resultados obtenidos por Nickell, en ella podemos observar los niveles de salario en el eje Y, mientras que en el eje X una representación de los datos obtenidos por los estudiantes en pruebas realizadas para obtener el nivel de sus habilidades académicas (Hanushek & Wößmann, Febrero 2017).



Note: Measure of inequality is the ratio of ninth decile to first decile in both cases; test performance refers to prose literacy in the International Adult Literacy Survey.

Source: Nickell (2004).

Figura 1 Nivel de salario con respecto a su formación académica (Hanushek & Ludger 2007)

La gráfica indica que el salario de un individuo se ve afectado por su nivel de habilidades académicas, que a su vez están directamente relacionadas con la calidad de la educación que recibe.

En el libro “Panorama educativo de México”, publicado en el 2012, el INNE pone a disposición, algunos indicadores del sistema educativo nacional.

Con estos indicadores, presenta un estudio acerca de los salarios por hora de los trabajadores adultos y jóvenes según su nivel de escolaridad. El resultado de este estudio puede ser observado en la siguiente gráfica:

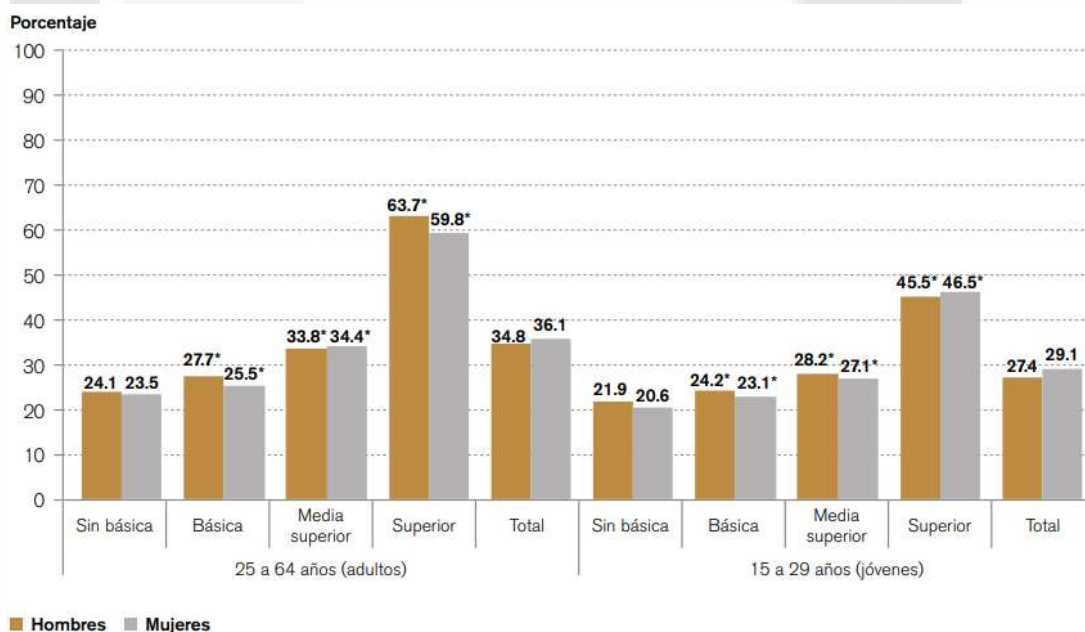


Figura 2 Salario por hora de los trabajadores adultos y jóvenes.

Este estudio, demuestra que el nivel de escolaridad afecta directamente en el ingreso económico de las personas. Pero, poseer el mismo grado de escolaridad no garantiza tener el mismo grado de conocimientos o habilidades, por lo tanto, no el mismo nivel de ingresos. Sin embargo, contar con una educación de calidad, si incrementa los conocimientos y habilidades de los alumnos; tomando en cuenta que esta calidad educativa está en función de las condiciones de vida de los alumnos, de la calidad de los docentes, selección apropiada de contenidos y mayor y mejor infraestructura.

2.3.3. El INEE como herramienta en la mejora educativa

En México desde inicios del siglo, el interés se ha centrado en indagar qué es lo que aprende la población escolar y cuáles son los aprendizajes básicos que forman parte en el conocimiento de los estudiantes. Ante esta situación, en nuestro país, al igual que en muchos otros, se estableció un sistema nacional de evaluación (el Instituto Nacional para la Evaluación de la Educación), el cual emplea pruebas diseñadas para tal efecto, además de cuestionarios de contexto que buscan encontrar las variables explicativas de los resultados obtenidos. En ese sentido, una de las funciones del INEE es evaluar lo que los estudiantes mexicanos aprenden del currículo nacional. Es decir, proveer información del logro educativo de los alumnos a lo largo del tiempo, así como de la calidad de los servicios que se ofrecen en la educación básica y media superior. Para el INEE el propósito fundamental de la evaluación del aprendizaje es proporcionar un conocimiento general del rendimiento académico de los estudiantes a niveles estatal y nacional, así como de los factores más importantes que influyen en éste (INEE, 2005).

2.3.4. La prueba EXCALE en la educación en México

Con el fin de conocer el desempeño educativo de los alumnos en México y los factores que influyen en este desempeño, el INEE planteó la necesidad de contar con instrumentos teóricos y técnicamente sólidos; idea que se consolidó con el desarrollo de una nueva generación de pruebas nacionales denominadas Exámenes de la Calidad y el Logro Educativos (Excale), mismas que se utilizaron por primera ocasión en junio de 2005.

Los resultados que produzcan los EXCALE deben permitir:

- a) Construir una visión general de lo que los estudiantes aprenden como resultado de su escolarización formal.
- b) Conocer los puntos fuertes y débiles del currículo nacional, atendiendo variables como grado y asignatura, modalidad educativa, género y estrato social.
- c) Establecer comparaciones del rendimiento escolar, así como de las tendencias de aprendizaje a lo largo del tiempo entre diversos grupos de alumnos y áreas temáticas.
- d) Mejorar la evaluación a gran escala de las disciplinas evaluadas.

Niveles, grados y dominios curriculares

El INNE considera que la evaluación del aprendizaje ha de centrarse en asignaturas propiamente instrumentales, tales como: matemáticas y español; además, de aquellas que cubren grandes áreas curriculares relacionadas con las Ciencias Naturales y Sociales, identificando en cada caso, la selección idónea de contenidos conforme al currículo nacional y al grado escolar a evaluar.

Respecto a los grados, se seleccionaron estratégicamente los terminales de cada ciclo escolar y los que representan momentos de inflexión en el desarrollo del aprendizaje en nuestro sistema educativo. De este modo, el INEE estableció las siguientes áreas temáticas para cada nivel y grado escolar.

Tabla 1 Niveles, Grados y asignaturas a evaluar (INEE, 2005)

Nivel	Grado	Dominio curricular
Prescolar	3°	Razonamiento Numérico
		Razonamiento Verbal
Primaria	3° Y 6°	Matemáticas
		Español
		Ciencias Naturales
		Ciencias Sociales
Secundaria	3°	Matemáticas
		Español
		Ciencias Naturales
		Ciencias Sociales
Bachillerato	3°	Matemáticas
		Español
		Ciencias Naturales
		Humanidades

2.3.4.1. El examen EXCALE a lo largo de los años

Para evaluar el aprendizaje, el INEE divide la educación básica en periodos de tres grados; por ese motivo, se aplican evaluaciones en tercero de preescolar: EXCALE 00, en tercero de primaria: EXCALE 03, en sexto de primaria: EXCALE 06 y en tercero de secundaria: EXCALE 09.

A partir de 2010, el INEE inició la aplicación de pruebas en la educación media superior a las que se les denomina EXCALE 124. Los EXCALE se han programado para aplicarse cada cuatro años al mismo grado escolar con el fin de poder dar seguimiento a una misma generación de estudiantes. Así, por ejemplo, la generación que estaba en tercero de primaria y que fue evaluada en 2006 (Excale 03) también fue evaluada en 2009 cuando cursaba el 6° grado (Excale 06), y se evalúa nuevamente en 2012, cuando los alumnos cursen 3° de secundaria (Excale 09) (Zamudio Mesa, Díaz Argüero, & Lepe García, 2012).

El orden de los grados en que se comenzaron a aplicar los EXCALE en el país, fue determinado por el INEE de acuerdo con su Plan General de Evaluación del Aprendizaje, y las decisiones tomadas al respecto, estuvieron sujetas a factores económicos y técnicos (INEE, 2005). El INEE ha aumentado de manera gradual los grados y asignaturas que se evalúan. En la Tabla 2 se muestran la asignatura y grado evaluados, así como el año de aplicación.

Tabla 2 Aplicaciones de los Excale por año (Zamudio Mesa, Díaz Argüero, & Lepe García, 2012)

	Ciclo escolar											
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
3° de preescolar Excale 00			E, M				E, M				E, M	
3° de primaria Excale 03		E, M CN, CS				E, M CN, FC y E				E, M CN, CS		
6° de primaria Excale 06	E, M		E, M		E, M CN, FC y E				E, M CN, CS			
3° de secundaria Excale 09	E, M			E, M CN, FC y E				E, M CN, CS				E, M CN, CS
Último grado de media superior Excale 12						E, CS						

2.3.4.2. Cuestionarios de contexto

Los resultados generados a partir de la aplicación de los EXCALE, permiten conocer lo que los estudiantes aprenden del currículo oficial; sin embargo, esta información no es suficiente para entender las diferencias en el logro académico. Para esto, el INEE desarrolla como estrategia paralela *los cuestionarios de contexto*, cuyo objetivo primordial es identificar los factores importantes asociados al rendimiento escolar.

Estos cuestionarios, están basados en el Modelo de Logro Educativo del INEE, en el que se plantea que el rendimiento académico de los estudiantes, se deriva de una compleja red de variables que interactúan entre sí para desencadenar efectos diferenciales en los distintos niveles del sistema educativo. En suma, este modelo propone que el logro académico está poderosamente influenciado por el esfuerzo coordinado de los alumnos, la familia, los docentes y las autoridades educativas (INEE, 2005).

Al igual que las pruebas de logro, los instrumentos de contexto se construyen siguiendo lineamientos y estándares apegados a una metodología robusta, en la que investigadores del INEE y un cuerpo colegiado de expertos externos determinan por consenso los dominios que deben incluir los cuestionarios que se aplican a los estudiantes evaluados, a sus maestros y al director de escuela.

La información del contexto debe ser suficiente y relevante para lograr conocer las diferencias en el aprendizaje (por género, grupos étnicos, nivel socioeconómico, situación laboral, modalidad o estrato educativo, y entidad federativa). Asimismo, permitirá identificar los factores personales, económicos y sociales que influyen en el aprendizaje (INEE, 2005).

El sistema integrado de información de los cuestionarios de contexto posibilitará, finalmente, dar sustento a las políticas públicas que consideran, en su hacer, las prácticas más eficaces de alumnos, padres de familia, docentes y directores.

2.4. Reconocimiento de Patrones

La selección de características forma parte del enfoque lógico-combinatorio del reconocimiento de patrones, razón por la cual se opta por comenzar a definirla y así, tener una visión de la posición en la que se trabajó para este documento de tesis.

A la fecha existen diferentes maneras de definir “reconocimiento de patrones” debido a que enmarcar una definición de una disciplina no es sencillo. Sin embargo, Ruíz Shulcloper hace una aproximación de la siguiente manera:

“Zona del conocimiento (de carácter interdisciplinario) que se ocupa del desarrollo de teorías, métodos, técnicas y dispositivos para la realización de procesos ingenieriles, computacionales y/o matemáticos, relacionados con objetos físicos y/o abstractos, que tienen el propósito de extraer información que permita establecer propiedades y/o vínculos de entre 21 un conjunto de dichos objetos sobre la base de los cuales se realiza una tarea de identificación o clasificación” (Ruíz Shulcloper, 2013)

En otras palabras, es el área multidisciplinaria de la ciencia, específicamente en inteligencia artificial, que se ocupa de procesos en ingeniería, computación y matemáticas relacionados con objetos, cuyo propósito es extraer información que permita establecer propiedades de o entre conjuntos de dichos objetos (Martínez Trinidad & Carrasco Ochoa, 2015). Así pues, el objetivo del reconocimiento de patrones es la clasificación y la recuperación de características únicas que identifiquen un sujeto de la misma especie o clase (Acosta, 2018).

Por sí mismo, el reconocimiento de patrones representa un desafío dentro de la inteligencia artificial debido al interés y alta demanda de aplicaciones especializadas en áreas como minería de datos, biometría, herramientas de toma de decisiones, entre otras.

En la actualidad, el reconocimiento de patrones es una de las líneas de investigación teórica y aplicada con alto auge que ha propiciado la existencia de diferentes enfoques entre los cuales se encuentra el estadístico-probabilístico, el sintáctico estructural, el lógico combinatorio, el neuronal y el asociativo (Martínez Trinidad & Carrasco Ochoa, 2015). Estos

enfoques no son necesariamente independientes y los problemas atacados pueden ser interpretados desde diferentes enfoques, como se puede observar en la Figura 3.

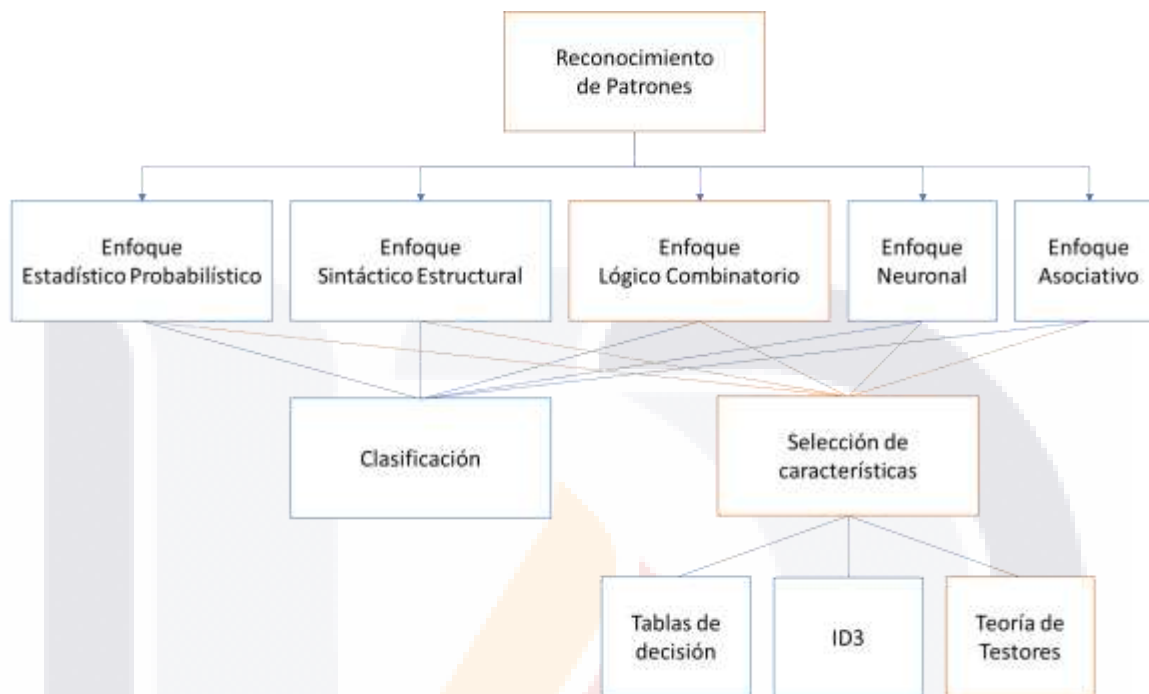


Figura 3 Enfoques del Reconocimiento de Patrones / Ubicación de la Selección de Características (Acosta, 2018)

2.4.1. Objetivos de los enfoques de reconocimiento de patrones

A continuación, se abordan brevemente el objetivo y las características de cada uno de los enfoques mencionados anteriormente descritos por Ariel Carrasco Ochoa y Francisco Martínez Trinidad en su artículo “Reconocimiento de Patrones” para la revista mexicana de inteligencia artificial Komputer Sapiens (Martínez Trinidad & Carrasco Ochoa, 2015):

- Enfoque estadístico-probabilístico
 - Basado en teoría de probabilidad y estadística.
 - Supone un conjunto de medidas numéricas con distribuciones de probabilidad conocidas o estimables, a partir de las cuales se hace el reconocimiento.
 - Históricamente, el primer enfoque abordado (Sánchez Garfias, Díaz de León Santiago, & Yáñez Márquez, 2003).

-
- Enfoque sintáctico-estructural:
 - Basado en teoría de Autómatas y lenguajes formales (Sánchez Garfias, Díaz de León Santiago, & Yáñez Márquez, 2003).
 - Estudio de objetos descritos como cadenas de símbolos, grafos, etc.
 - Enfocado en la búsqueda de relaciones estructurales de los objetos de estudio.
 - Redes Neuronales:
 - Resolución de problemas por medio de modelo matemáticos de las redes neurológicas biológicas (Sánchez Garfias, Díaz de León Santiago, & Yáñez Márquez, 2003).
 - Estas redes se entrenan para dar cierta respuesta cuando se presentan determinados valores en sus entradas.
 - Una red neuronal puede dar una respuesta similar cuando recibe entradas parecidas a la usadas en su entrenamiento.
 - Enfoque asociativo (Yáñez, 2002):
 - Utiliza modelos de memorias asociativas para crear clasificadores robustos.
 - El propósito de una memoria asociativa es recuperar patrones completos a partir de patrones de entrada con posible ruido.
 - Creado en el Centro de Investigación en Computación del IPN en 2002. 23
 - Lógico-Combinatorio:
 - Alternativa a los enfoques anteriores.
 - Basado en la idea de que el modelo de un problema debe ser lo más cercano a la realidad del mismo.
 - Los atributos que describen a los objetos de estudio son tratados cuidadosamente para no resultar en operaciones antinaturales.
 - Acepta atributos cualitativos y cuantitativos e incluso con ausencia de información.

2.4.2. Problemas que ataca el Reconocimiento de Patrones

En este apartado, se busca describir las características de los problemas que el reconocimiento de patrones puede resolver por medio de sus diferentes enfoques entre los cuales se encuentra la selección de características, que determina el conjunto más adecuado para describir objetos; la clasificación supervisada, para la clasificación de objetos por medio de una muestra ya clasificada; y la clasificación no supervisada, donde la clasificación no requiere información de clasificación previa de los objetos e incluso, sin definición de clases (Martínez Trinidad & Carrasco Ochoa, 2015).

- Selección de Características (Sánchez Garfías, Díaz de León Santiago, & Yáñez Márquez, 2003):
 - Determina el conjunto de características/variables más adecuando para describir objetos.
 - Por tanto, remueve variables poco relevantes (ruido o distractores).
 - Mejoramiento del desempeño de los clasificadores.
 - Representaciones estables de los objetos.
 - Facilidad de visualización y comprensión de datos.
- Clasificación (Martínez Trinidad & Carrasco Ochoa, 2015):
 - Asignación de un objeto/fenómeno a una de las categorías.
 - Clases que se especifiquen dependiendo de las características que comparta con una clase definida o ampliamente utilizada en el reconocimiento de patrones.
- Clasificación Supervisada (Grupo de Tecnologías Avanzadas en Computación, 2008):
 - Clasificación de objetos por medio de una muestra ya clasificada (conjunto de entrenamiento).
 - Compuesta de dos fases: entrenamiento y clasificación.
 - En la fase de entrenamiento se cuenta propiamente con el conjunto de entrenamiento y uno adicional para validación. Ambos constituyen un modelo de clasificación.

-
-
- Clasificación no Supervisada (Grupo de Tecnologías Avanzadas en Computación, 2008):
 - No cuenta con conocimiento de clasificación previa.
 - Algunas veces, no hay definición previa de clases.
 - También conocida como *clustering*.
 - Se basa en el descubrimiento de grupos de objetos cuyas características similares de por resultado la separación en clases diferentes.

2.5. Testores y Testores Típicos

Existen diferentes herramientas o metodologías para llevar a cabo la selección de características (Ver Figura 3), de las cuales, en este documento, se aborda la teoría de testores formulada como dirección científica de Cibernética Matemática en los años 60 en la Unión de Repúblicas Socialistas Soviéticas (URSS), cuyo origen está vinculado con el uso de lógica matemática para localizar fallas en circuitos electrónicos que realizan funciones booleanas (Ruíz Shulcloper, Alba Cabrera, & Lazo Cortés, 1995).

El enfoque de la teoría de testores a utilizar, es el de Dimitriev, Zhuravlev y Krendeleiev para problemas clásicos del reconocimiento de patrones (clasificación con aprendizaje y selección de características) creado en 1965 donde “*las clases son conjuntos disjuntos, el criterio de comparación entre rasgos es booleano y el criterio de semejanza entre objetos asume que dos objetos son diferentes si al menos uno de sus rasgos también lo es*” (Lias-Rodríguez & Pons-Porrata).

Uno de los conceptos que define ésta teoría es el de testor, el cual es un subconjunto de características que distingue objetos de diferentes clases [3], es decir, ningún objeto perteneciente a una clase T_0 se confunde con algún objeto de la clase T_1 mirados a través de los valores en sus características (Ochoa Somuano, 2005).

De acuerdo a Shulcloper, la definición de testor se puede extender a más de dos clases como lo hace Zhuravlev, de manera que, dado un subconjunto de columnas $\tau = \{i_1, \dots, i_s\}$ de columnas de la tabla T y sus rasgos $(a_{i_1}, \dots, a_{i_s})$, se considera testor para $(T_0, \dots, T_r)=T$,

si después de eliminar de T todas las columnas de τ no existe alguna fila en T_0 igual a una en cualquiera del resto de las r clases en T (Ruíz Shucloper, Alba Cabrera, & Lazo Cortés, 1995).

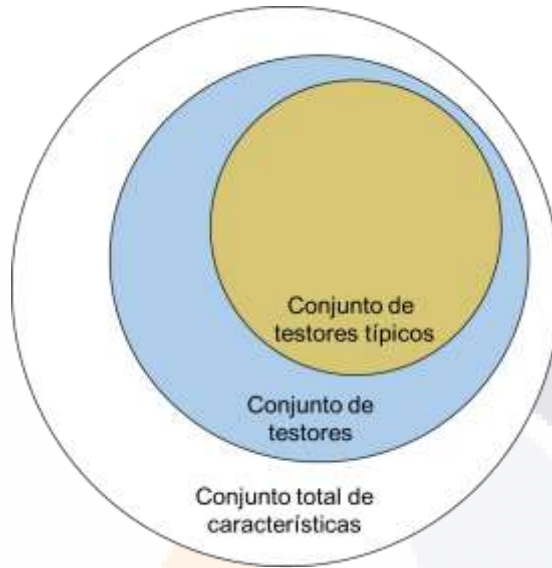


Figura 4 Conjuntos de Características (Acosta, 2018)

Dentro del conjunto de testores se encuentran los llamados testores típicos o testores irreducibles (Ver Figura 4), los cuales son testores que si perdieran cualquiera de sus características perderían a su vez el estado de testor (Ruíz Shucloper, Alba Cabrera, & Lazo Cortés, 1995). Por tanto, un testor típico es el subconjunto de características mínimo necesario para diferenciar objetos de diferentes clases.

Así pues, la importancia del conjunto de testores radica propiamente en la reducción del espacio de representación de objetos (selección de características), como apoyo en la evaluación de peso informacional de características y como apoyo a sistemas de clasificación (Santiesteban Alganza & Pons Porrata, 2003).

3. METODOLOGÍA

3.1. Actividades Principales

En la presente investigación, se realizó un estudio de los factores que determinan el desempeño educativo en la prueba EXCALE, este problema abordó desde un enfoque lógico-combinatorio de reconocimiento de patrones como parte del área de la inteligencia artificial partiendo de información alojada públicamente en el sitio oficial del INEE.

La metodología implementada en la presente investigación cuenta con las fases mostradas en la siguiente figura:



Figura 5 Metodología utilizada.

Las fases de la metodología mostrada en la figura 5, así como la descripción de sus componentes, se describen a continuación:

Recolección de información

Esta sección está enfocada a la obtención de la información cruda con la que se trabajó. La información se encuentra alojada públicamente en la página oficial del Instituto Nacional Para la Evaluación de la Educación. El INEE resguarda resultados de todas las pruebas de los Exámenes de la Calidad y Logro Educativo (EXCALE) que ha realizado desde el 2005, así como los correspondientes cuestionarios de contexto; para esta investigación se trabajó con la prueba EXCALE_03_2014, específicamente con la asignatura de matemáticas.

Preprocesamiento

Consiste en el procesamiento previo de los datos obtenidos en la sección anterior, con el fin de elegir los datos que son necesarios para el estudio y eliminar aquellos que no tengan relevancia, de igual manera, organizar la información en los diferentes grupos de estudio y por último generar una base de datos que sirva de entrada para el algoritmo de testores típicos.

Implementación del algoritmo de testores típicos

En esta sección se explica el algoritmo de testores típicos y se desarrolla una adaptación para los diferentes requerimientos de la base de datos creada con anterioridad, esta adaptación depende del número de variables que contenga la base de datos.

Construcción de herramienta para validación

Para esta sección, se explica el desarrollo, la estructura y el funcionamiento de un modelo de clasificación creado en KNIME, que permitió evaluar la eficacia de los testores, de manera que se logre saber en qué porcentaje es posible clasificar a los sustentantes en base a las características que el testor representa.

Análisis de Resultados

Con la implementación del algoritmo se obtienen los diferentes testores típicos existentes, de ellos se obtiene el peso informacional de cada variable, con lo cual se determinan los factores que determinan el desempeño educativo en la prueba EXCALE. Por último, se lleva a cabo una validación de los factores que inciden en el desempeño educativo, mediante el uso de KNIME, un software computacional para la minería de datos

Conclusiones

Se muestran a manera de resumen los resultados obtenidos que han sido validados, así como las contribuciones de este trabajo.

3.2. Recolección de información

El Instituto Nacional Para la Evaluación de la Educación en México, lleva años desarrollando evaluaciones y además de evaluar los resultados obtenidos en dichas evaluaciones, tiene el compromiso de almacenar digitalmente de manera segura esta información. Además de su resguardo, el INNE pone a disposición esta información para ayudar a investigadores nacionales e internacionales a llevar a cabo indagaciones que puedan generar nuevo conocimiento.

Estos resultados de dichas evaluaciones se ofrecen de manera pública en el sitio oficial del INNE, en el apartado de EXCALE, en la URL <https://www.inee.edu.mx/evaluaciones/planea/excale/>. La presentación de estos resultados se encuentra de manera jerárquica, comenzando por las diferentes presentaciones de los EXCALE, y a continuación se muestran los diferentes archivos correspondientes a la prueba seleccionada, entre estos archivos, se encuentran los resultados de la evaluación en las diferentes materias aplicadas en la prueba, así como los resultados de los cuestionarios de contexto para alumnos, maestros y administrativos. Además de los resultados de contexto y de desempeño; así mismo, se presentan los formatos de los cuestionarios y evaluaciones aplicados.

Para esta investigación se seleccionó la prueba “EXCALE 03_2014”, que corresponde a la evaluación de los alumnos del Tercer grado de Primaria al finalizar el

ciclo escolar 2013-2014, en específico los resultados de la evaluación en Matemáticas. Incluyendo todas las pruebas de las diferentes asignaturas, se administraron las pruebas EXCALE a un total de 31933 alumnos, de 3552 escuelas en todo el país para evaluar su desempeño en las asignaturas de:

- Español
- Matemáticas
- Formación cívica y ética
- Ciencias naturales

Respecto de los cuestionarios, éstos abordan aspectos sobre el contexto en el que se desenvuelven alumnos, docentes y directores.

3.3. Preprocesamiento

Debido a que los datos disponibles en el sitio oficial del INNE se encuentran en un formato poco funcional para la manipulación de los datos, lo primero que se realizó fue migrar la información necesaria para la investigación a un formato que facilitara dicha manipulación, debido a la experiencia y a la facilidad de procesamiento, se seleccionó el uso de una base de datos en MySQL que se utilizó para el manejo de los datos. El formato inicial en que se presentan los datos, es un archivo comprimido que contiene archivos necesarios para su lectura en el sistema de análisis estadístico SPSS. Estos archivos, incluyen las especificaciones y/o estructuras en que vienen organizados los datos, así como una descripción de cada uno de los campos, la manera en que fueron evaluados, además de la forma en que se encuentran representados.

Por otro lado, el archivo comprimido contiene un archivo con extensión .txt que contiene los datos brutos de los resultados separando cada registro en un renglón y las columnas por posicionamiento en el archivo. Al abrir los archivos de características en el software SPSS es posible manipular de manera práctica las evaluaciones dispuestas por el INEE.

Afortunadamente, existen herramientas dentro del Software SPSS que propician la fácil exportación de información a bases de datos MySQL.

Es necesario mencionar que no toda la información presentada por el INNE fue considerada para llevar a cabo esta investigación, para el caso de los cuestionarios de contexto existían campos que eran repetitivos en cuanto la información que representan, y con respecto de los resultados de la evaluación, contiene campos que son referentes a la recolección de datos y no al resultado obtenido en la prueba.

3.3.1. Transformación de variables de trabajo en la base de datos de la evaluación de contexto

Para reestructurar nuestra información es necesario evaluar primero la base de datos que contienen los resultados de la evaluación de contexto de los sustentantes de la prueba EXCALE. La base de datos cuenta con diferentes campos, estos campos deben ser agrupados en diferentes rubros para generar los constructos que al final sirven de entrada para el algoritmo de testores típicos. A cada valor de cada campo de las diferentes agrupaciones se le debe asignar una ponderación individual de acuerdo con la respuesta en esa sección, en conjunto, las ponderaciones de cada agrupación deben generar el valor que representará a ese constructo para cada uno de los registros. A continuación, se muestran los diferentes constructos con sus campos o variables que los componen.

Tabla 3 Constructo Edad Normativa (CI)

Pregunta	Respuesta	Ponderación
Edad normativa. Supone el ingreso al primero de primaria con 6 años cumplidos (o menos) al primero de septiembre	1 - Edad normativa	1 - 1
	2 - Extra - edad	2 - 2
	9 - Respuesta omitida	9 - Respuesta omitida

Tabla 4 Constructo Vivienda (C2)

Pregunta	Respuesta	Ponderación
¿De qué material es la mayor parte del piso que está dentro de tu casa?	1 - De tierra 2 - De cemento o firme 3 - De mosaico, alfombra, loseta, duela u otro recubrimiento 8 - Respuesta múltiple 9 - Respuesta omitida /	1 - 0.33 2 - 0.66 3 - 1 8 - Respuesta múltiple 9 - Respuesta omitida /
¿Para tener agua dentro de tu casa (para cocinar, bañarte, lavar trastos, etc.), debes salir a buscarla?	1 - Nunca o casi nunca 2 - Pocas veces 3 - Muchas veces 4 - Siempre o casi siempre 8 - Respuesta múltiple 9 - Respuesta omitida /	1 - 1 2 - 0.75 3 - 0.5 4 - 0.25 8 - Respuesta múltiple 9 - Respuesta omitida /
Al baño, escusado o letrina de tu casa...	3 - hay que jalarle la palanca para que se vaya el agua 2 - hay que echarle agua con una cubeta 1 - no se le puede echar agua 8 - Respuesta múltiple 9 - Respuesta omitida /	3 - 1 2 - 0.66 1 - 0.33 8 - Respuesta múltiple 9 - Respuesta omitida /
¿Cuántos cuartos de baño hay en la casa donde vives?	1 - Ninguno 2 - 1 3 - 2 4 - 3 o más 8 - Respuesta múltiple 9 - Respuesta omitida /	1 - 0.25 2 - 0.5 3 - 0.75 4 - 1 8 - Respuesta múltiple 9 - Respuesta omitida /
¿Cuántos focos hay en la casa donde vives (incluyendo los de las lámparas)?	1 - 1 o 2 2 - 3 o 4 3 - 5 o 6 4 - 7 u 8 5 - 9 o 10 6 - Entre 11 y 15 7 - Entre 16 y 25	1 - 0.2 2 - 0.4 3 - 0.6 4 - 0.8 5 - 1 6 - Entre 11 y 15 7 - Entre 16 y 25
¿En la casa donde vives, duermen tres personas o más en un mismo cuarto?	1 - Sí 0 - No 8 - Respuesta múltiple 9 - Respuesta omitida /	1 - 0.5 0 - 1 8 - Respuesta múltiple 9 - Respuesta omitida /

Tabla 5 Constructo Situación Económica (C3)

Pregunta	Respuesta	Ponderación
¿Has recibido o recibes beca del Programa Oportunidades?	1 - Sí	1 - 0.33
	0 - No	0 - 1
	2 - No sé	2 - 0.66
	8 - Respuesta múltiple	8 - Respuesta múltiple
	9 - Respuesta omitida /	9 - Respuesta omitida /
La última vez que saliste de vacaciones con tu familia, ¿a dónde fueron?	1 - Nunca hemos salido de vacaciones	1 - 0.25
	2 - A algún lugar dentro del estado donde vivo	2 - 0.5
	3 - A algún lugar en otro estado del país	3 - 0.75
	4 - A otro país	4 - 1
	8 - Respuesta múltiple	8 - Respuesta múltiple
9 - Respuesta omitida /	9 - Respuesta omitida /	
¿Tienen internet en tu casa?	1 - Sí 0 - No 8 - Respuesta múltiple 9 - Respuesta omitida /	1 - 1 0 - 0.5 8 - Respuesta múltiple 9 - Respuesta omitida /
¿Tienen televisión de paga en tu casa (por cable o satelital)?		
¿Tienen teléfono en tu casa?		
¿Tienen automóvil o camioneta en tu casa?		
¿Tienen computadora en tu casa?		
¿Tienen lavadora de ropa en tu casa?		
¿Tienen horno de microondas en tu casa?		
¿Tienen videocasetera o DVD en tu casa?		

¿Cuántas personas viven en tu casa?	1 - 2 personas 2 - 3 personas 3 - 4 personas 4 - 5 personas 5 - 6 personas 6 - 7 personas 7 - 8 personas 8 - 9 personas 9 - 10 o más personas 98 - Respuesta múltiple 99 - Respuesta omitida /	1 - 2 personas 1 2 - 3 personas 1 3 - 4 personas 0.75 4 - 5 personas 0.5 5 - 6 personas 0.4 6 - 7 personas 0.3 7 - 8 personas 0.2 8 - 9 personas 0.1 9 - 10 o más personas 0.1 98 - Respuesta múltiple 99 - Respuesta omitida /
¿Cuántas personas de las que viven en tu casa trabajan y reciben un pago por ello?	1 - 1 persona 2 - 2 personas 3 - 3 personas 4 - 4 personas 5 - 5 personas 6 - 6 personas 7 - 7 o más personas 98 - Respuesta múltiple 99 - Respuesta omitida /	1 - 1 persona 0.5 2 - 2 personas 0.6 3 - 3 personas 0.7 4 - 4 personas 1 5 - 5 personas 1 6 - 6 personas 1 7 - 7 o más personas 1 98 - Respuesta múltiple 99 - Respuesta omitida /
Además de asistir a la escuela, ¿trabajas?	1 - No trabajo 2 - Sí trabajo, pero no me pagan 3 - Sí trabajo y me dan cosas a cambio 4 - Sí trabajo y me pagan 8 - Respuesta múltiple 9 - Respuesta omitida /	1 -1 2 -0.75 3 -0.5 4 -0.5 8 - Respuesta múltiple 9 - Respuesta omitida /

Tabla 6 Constructo Contexto Padres (C4)

Pregunta	Respuesta	Ponderación
¿Tu mamá sabe leer y escribir?	1 - Sí 0 - No	1 - Sí 1 0 - No 0.25
¿Tu mamá fue a la preparatoria o al bachillerato?	2 - No sé 3 - No tengo mamá 8 - Respuesta múltiple	2 - No sé 0.75 3 - No tengo mamá 0.25 8 - Respuesta múltiple
¿Tu mamá fue a la universidad?	9 - Respuesta omitida /	9 - Respuesta omitida /
¿Tu mamá trabaja y recibe un pago por ello?		
¿Tu papá sabe leer y escribir?	1 - Sí 0 - No	1 - Sí 1 0 - No 0.25
¿Tu papá fue a la preparatoria o al bachillerato?	2 - No sé 3 - No tengo papá 8 - Respuesta múltiple	2 - No sé 0.75 3 - No tengo papá 0.25 8 - Respuesta múltiple
¿Tu papá fue a la universidad?	9 - Respuesta omitida /	9 - Respuesta omitida /
¿Tu papá trabaja y recibe un pago por ello?		

Tabla 7 Constructo Futuro Intelectual (C5)

Pregunta	Respuesta	Ponderación
¿Hasta qué grado les gustaría a tus padres que tú estudiaras?	1 - Primaria 2 - Secundaria 3 - Bachillerato o preparatoria 4 - Carrera universitaria 5 - No sé 8 - Respuesta múltiple 9 - Respuesta omitida /	1 -0.2 2 -0.6 3 -0.8 4 -1 5 -0.4 8 - Respuesta múltiple 9 - Respuesta omitida /
A ti, ¿hasta qué grado te gustaría estudiar?	1 - Primaria 2 - Secundaria 3 - Bachillerato o preparatoria 4 - Carrera universitaria 8 - Respuesta múltiple 9 - Respuesta omitida /	1 -0.25 2 -0.5 3 -0.75 4 -1 8 - Respuesta múltiple 9 - Respuesta omitida /

Tabla 8 Constructo Entorno Casa Escuela (C6)

Pregunta	Respuesta	Ponderación
¿Cuánto tiempo tardas en llegar de tu casa a la escuela?	1 - Menos de media hora	1 -1
	2 - Entre 30 y 60 minutos	2 -0.75
	3 - Entre 1 y 2 horas	3 -0.5
	4 - Más de 2 horas	4 -0.25
	8 - Respuesta múltiple	8 - Respuesta múltiple
	9 - Respuesta omitida /	9 - Respuesta omitida /
La mayoría de las veces, ¿qué transporte utilizas para llegar a la escuela?	1 - Me voy caminando	1 - Me voy caminando 0.2
	2 - Bicicleta, patineta, patines, etc.	2 - Bicicleta, patineta, patines, etc. 0.4
	3 - Transporte público (camión, metro, combi, micro, autobús, colectivo, taxi, etc.)	3 - Transporte público (camión, metro, combi, micro, autobús, colectivo, taxi, etc.) 0.5
	4 - El camión de la escuela, auto, moto o camioneta	4 - El camión de la escuela, auto, moto o camioneta 1
	5 - Uso un transporte distinto a todos los anteriores	5 - Uso un transporte distinto a todos los anteriores 0.6
	8 - Respuesta múltiple	8 - Respuesta múltiple
	9 - Respuesta omitida /	9 - Respuesta omitida /

Tabla 9 Constructo Lengua Indígena (C7)

Pregunta	Respuesta	Ponderación
¿Alguno de tus padres o de tus abuelos habla una lengua indígena (náhuatl, maya, purépecha, mixteco, otomí u otra)?	1 - Sí	1 - 0.5
	0 - No	0 - 1
	8 - Respuesta múltiple	8 - Respuesta múltiple
	9 - Respuesta omitida /	9 - Respuesta omitida /

Tabla 10 Constructo "condiciones de estudio" (C8)

Pregunta	Respuesta	Ponderación
¿Alguien en tu casa está al pendiente de que hagas la tarea?	4 - Sí, todos los días 3 - Algunos días sí y otros no. 2 - Casi nunca 1 - No, nunca 8 - Respuesta múltiple 9 - Respuesta omitida/	4 - 1 3 - 0.75 2 - 0.3 1 - 0.1 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Alguien en tu casa te ayuda a resolver tus dudas de la escuela o a hacer la tarea?	4 - Sí, siempre que lo necesito 3 - A veces sí y otras no 2 - Casi nunca 1 - No, nunca 8 - Respuesta múltiple 9 - Respuesta omitida/	4 - 1 3 - 0.75 2 - 0.5 1 - 0.25 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Para hacer la tarea o estudiar en casa Sí tienes un lugar silencioso?	1 - Sí 0 - No	1 - 1 0 - 0.5
¿Para hacer la tarea o estudiar en casa tienes una mesa y una silla?	8 - Respuesta múltiple 9 - Respuesta omitida/	8 - Respuesta múltiple 9 - Respuesta omitida/
¿Para hacer la tarea o estudiar en casa tienes un lugar con luz suficiente?		
¿En casa tienes un espacio para guardar tus cosas de la escuela?		
En tu casa, ¿cuántos libros hay aparte de los que usas en la escuela? (no incluyas revistas ni periódicos)	1 - Ninguno 2 - Como 10 libros 3 - Como 25 libros 4 - Como 50 libros 5 - Como 100 libros 6 - Más de 100 libros 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 0.17 2 - 0.33 3 - 0.5 4 - 0.75 5 - 1 6 - 1 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Tienes en tu casa papel y lápices, además de los que usas para la escuela?	1 - Sí 0 - No	1 - 1 0 - 0.75
¿Tienes en tu casa colores o pinturas, tijeras o pegamento para tu uso?	8 - Respuesta múltiple 9 - Respuesta omitida/	8 - Respuesta múltiple 9 - Respuesta omitida/

¿Tienes en tu casa juegos de mesa (lotería, memoria, rompecabezas, ajedrez, dominó, etc.)?		
¿Tienes en tu casa una computadora que tú puedas utilizar?		
¿Tienes en tu casa libros o materiales que te ayuden a resolver tus dudas de la escuela (enciclopedias, internet, etc.)?		
¿Has tomado alguna clase fuera del horario de la escuela (de idiomas, música, deportes, danza, computación, etc.)?	1 - Sí 0 - No 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 1 0 - 0.75 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Algún adulto de tu casa lee libros, periódicos, revistas, etc.?	1 - Nunca o casi nunca 2 - Pocas veces 3 - Muchas veces 4 - Siempre o casi siempre 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 0.25 2 - 0.5 3 - 1 4 - 1 8 - Respuesta múltiple 9 - Respuesta omitida/
A ti, ¿te gusta leer?	4 - Mucho 3 - Más o menos 2 - Poco 1 - Casi nada 8 - Respuesta múltiple 9 - Respuesta omitida/	4 - 1 3 - 0.75 2 - 0.5 1 - 0.25 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Qué tan seguido lees un libro o revista con algún miembro de tu familia?	1 - Nunca o casi nunca 2 - Pocas veces	1 - 0.25 2 - 0.5
¿Qué tan seguido van al cine o al teatro?	3 - Muchas veces	3 - 0.75
¿Qué tan seguido van a museos?	4 - Siempre o casi siempre	4 - 1
¿Qué tan seguido escuchan música?	8 - Respuesta múltiple	8 - Respuesta múltiple
¿Qué tan seguido van a eventos musicales o de danza?	9 - Respuesta omitida/	9 - Respuesta omitida/
¿Qué tan seguido utilizan internet?		

Tabla 11 Constructo Actividades Extracurriculares (C9)

Pregunta	Respuesta	Ponderación
¿Has tomado alguna clase fuera del horario de la escuela (de idiomas, música, deportes, danza, computación, etc.)?	1 - Sí 0 - No 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 1 0 - 0.75 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Algún adulto de tu casa lee libros, periódicos, revistas, etc.?	1 - Nunca o casi nunca 2 - Pocas veces 3 - Muchas veces 4 - Siempre o casi siempre 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 0.25 2 - 0.5 3 - 1 4 - 1 8 - Respuesta múltiple 9 - Respuesta omitida/
A ti, ¿te gusta leer?	4 - Mucho 3 - Más o menos 2 - Poco 1 - Casi nada 8 - Respuesta múltiple 9 - Respuesta omitida/	4 - 1 3 - 0.75 2 - 0.5 1 - 0.25 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Qué tan seguido lees un libro o revista con algún miembro de tu familia?	1 - Nunca o casi nunca 2 - Pocas veces	1 - 0.25 2 - 0.5
¿Qué tan seguido van al cine o al teatro?	3 - Muchas veces	3 - 0.75
¿Qué tan seguido van a museos?	4 - Siempre o casi siempre	4 - 1
¿Qué tan seguido escuchan música?	8 - Respuesta múltiple	8 - Respuesta múltiple
¿Qué tan seguido van a eventos musicales o de danza?	9 - Respuesta omitida/	9 - Respuesta omitida/
¿Qué tan seguido utilizan internet?		

Tabla 12 Constructo Antecedentes Académicos (C10)

Pregunta	Respuesta	Ponderación
¿Cuántos años fuiste al preescolar, kínder o jardín de niños?	1 - Ninguno 2 - 1 año 3 - 2 años 4 - 3 años 5 - 4 años o más 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 0.25 2 - 0.5 3 - 0.75 4 - 1 5 - 1 8 - Respuesta múltiple 9 - Respuesta omitida/
Desde que entraste a la primaria, ¿has repetido algún grado?	1 - Ninguno 2 - He repetido 1 grado 3 - He repetido 2 grados o más 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 1 2 - 0.66 3 - 0.33 8 - Respuesta múltiple 9 - Respuesta omitida/
Desde que entraste a la primaria, ¿en cuántas escuelas has estudiado?	1 - Solamente en esta escuela 2 - En 2 escuelas 3 - En 3 escuelas o más 8 - Respuesta múltiple 9 - Respuesta omitida/	1 - 1 2 - 0.66 3 - 0.33 8 - Respuesta múltiple 9 - Respuesta omitida/
Desde que entraste a la primaria, ¿te has cambiado de pueblo o de ciudad?	0 - No 1 - Sí, una vez 2 - Sí, más de una vez 8 - Respuesta múltiple 9 - Respuesta omitida/	0 - 1 1 - 0.66 2 - 0.33 8 - Respuesta múltiple 9 - Respuesta omitida/
¿Cuántos años tienes viviendo en el mismo pueblo o ciudad donde vives ahora?	1 - 1 año0.33 2 - 2 años0.33 3 - 3 años0.66 4 - 4 años0.66 5 - 5 años1 6 - 6 años1 7 - 7 años1 8 - 8 años1 9 - 9 años1 10 - 10 años1 11 - No sé0.5 98 - Respuesta múltiple 99 - Respuesta omitida	1 - 1 año0.33 2 - 2 años0.33 3 - 3 años0.66 4 - 4 años0.66 5 - 5 años1 6 - 6 años1 7 - 7 años1 8 - 8 años1 9 - 9 años1 10 - 10 años1 11 - No sé0.5 98 - Respuesta múltiple 99 - Respuesta omitida

Para lograr generar el valor que representará a cada uno de los constructos y sus características, en cada alumno, fue necesario realizar una suma entre dichas características, al concluir este proceso se obtuvo una tabla como la siguiente:

Tabla 13 Ejemplo de valores obtenidos por constructo

Alumno	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos
1	1	11	1.43	7.25	2	1.2	1	10.83	5	4
2	2	12.11	2.5	7.25	1.75	2	1	10.83	7.5	4.33
3	2	10.17	3	5	0.5	1.2	1	11.25	5.25	4.5
4	1	13.33	2.6	8	2	1.2	0.5	10.92	6.75	5
5	1	12.32	1.68	7	2	0.7	0.5	11	5	5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	nC1	nC2	nC3	nC4	nC5	nC6	nC7	nC8	nC9	nC10

3.3.2. Transformación de las variables de trabajo de la base de datos de resultados en la prueba de matemáticas

Conociendo con anterioridad las diferentes características (constructos) que definen a nuestros individuos, en este caso alumnos que presentaron la prueba “EXCALE 03_2014”, corresponde ahora definir las clases a las cuales pertenecen cada uno de los alumnos, para esto, con el uso de sus resultados en la prueba, se calificó su desempeño en base a los resultados obtenidos en el logro académico de los alumnos. En la siguiente tabla, se presentan los diferentes logros académicos que se evaluaron en la prueba, y que se utilizaron para la calificación y posterior clasificación de nuestros datos.

Tabla 14 Logros académicos evaluados en la prueba EXCALE

	Logro Académico
1	Reconocer la operación de suma que resuelve un problema aditivo.
2	Identificar la resta que resuelve un problema aditivo.
3	Resolver problemas aditivos que incluyan dos operaciones combinadas, con base en la información dada explícitamente en un portador.
4	Resolver problemas de multiplicación con datos dados en un cartel.
5	Resolver problemas de reparto que involucren una división sin el apoyo de una ilustración.
6	Identificar fracciones equivalentes mayores que la unidad.
7	Resolver problemas de reparto cuya solución implique un cociente con información dada en portadores.
8	Identificar la ubicación de un objeto por su longitud en un conjunto ordenado de objetos.
9	Reconocer longitudes o distancias que es conveniente medir en metros.
10	Resolver problemas de reparto cuyo resultado sea una fracción impropia.
11	Identificar números no consecutivos en una sucesión decreciente con progresión aritmética.
12	Resolver problemas aditivos de combinación que implique reunir dos conjuntos y la incógnita es el resultado.
13	Resolver problemas aditivos de cambio en los que la incógnita es el operador.
14	Resolver problemas aditivos de comparación donde la incógnita es la cantidad mayor.
15	Resolver problemas de comparación donde la incógnita es la diferencia.
16	Resolver una multiplicación de un múltiplo de 10 menor que 100 por un número de una cifra.
17	Resolver problemas de agrupamiento con datos en un cartel cuya solución implique una división.
18	Identificar figuras no consecutivas que completan una sucesión creciente con progresión aritmética.
19	Resolver problemas de agrupamiento con divisores menores que 10, usando una ilustración, cuya solución implique una división.
20	Resolver problemas que impliquen calcular la hora en que finaliza una actividad, dada la hora de inicio y la duración.
21	Resolver un problema en donde se calcula la hora en que finaliza una actividad, a partir de la hora marcada en un reloj de manecillas.
22	Ordenar de mayor a menor números naturales de cuatro cifras.
23	Identificar cómo se nombran, los números escritos con cuatro cifras con ceros intermedios.
24	Resolver problemas de transformación con la incógnita en el estado final.
25	Resolver problemas aditivos de comparación en los que la incógnita es la cantidad menor.
26	Resolución de problemas aditivos de igualación donde la incógnita sea la diferencia.
27	Resolver problemas de reparto que involucren una división con el apoyo de una ilustración.

28	Restar números naturales de tres cifras con transformación y ceros en el minuendo.
29	Resolver problemas que implican multiplicar un número de dos cifras por uno de una, con transformación.
30	Reconocer la división como la operación que puede resolver un problema multiplicativo.
31	Resolver problemas que implican calcular la duración de una actividad, a partir de la hora marcada en un reloj de manecillas.
32	Comparar las longitudes de distintos objetos.
33	Identificar las características de una figura geométrica dada su imagen.
34	Identificar cómo se escriben un número de cuatro cifras sin ceros intermedios dado su nombre con letra.
35	Identificar fracciones equivalentes menores que la unidad.
36	Resolver problemas aditivos de cambio donde la incógnita es la cantidad inicial.
37	Identificar tres números consecutivos que continúan en una sucesión decreciente con progresión aritmética.
38	Resolver problemas aditivos de cambio en los que la incógnita está en el operador.
39	Sumar números naturales de tres cifras sin transformación.
40	Resolver problemas aditivos que implican tres operaciones a partir de información dada en portadores.
41	Identificar las figuras que continúan una sucesión creciente con progresión aritmética.
42	Resolver problemas de multiplicación con factores de un dígito.
43	Identificar la suma iterada que resuelve un problema multiplicativo.
44	Multiplicar dos números de una cifra.
45	Resolver problemas que impliquen calcular la duración de una actividad dada la hora de inicio y de término.
46	Ordenar de menor a mayor tres objetos por su longitud.
47	Resolver problemas de reparto cuyo resultado sea una fracción propia.
48	Identificar tres números consecutivos que continúan en una sucesión creciente con progresión aritmética.
49	Identificar dos números no consecutivos en una sucesión creciente con progresión aritmética.
50	Resolver problemas aditivos de combinación donde la incógnita es uno de los sumandos.
51	Resolver problemas aditivos de cambio en los que la incógnita es el estado final.
52	Restar números naturales de tres cifras sin transformación con ceros en el sustraendo.
53	Restar números naturales de tres cifras con transformación y sin ceros en el minuendo.
54	Resolver problemas de multiplicación con factores de dos cifras por uno de una, sin transformación.
55	Reconocer una composición formada por figuras geométricas idéntica a otra.
56	Resolver un problema de agrupamiento que implique una división sin apoyo de una ilustración.

57	Identificar un triángulo dadas sus características geométricas tales como número, tamaño y forma de lados.
58	Identificar cómo se escribe con cifras un número de cuatro cifras con cero intermedios, dado su nombre con letra.
59	Comparar números naturales de cuatro cifras.
60	Resolver problemas aditivos de cambio con la incógnita en el estado inicial.
61	Resolver problemas aditivos de igualación donde la incógnita es la cantidad mayor.
62	Resolver problemas aditivos de igualación donde la incógnita es la cantidad menor.
63	Identificar la operación de multiplicación que permite resolver un problema.
64	Calcular el factor de una multiplicación cuando se da el resultado del producto y el otro factor.
65	Identificar la hora marcada en un reloj de manecillas.
66	Resolver problemas que impliquen calcular la hora de inicio de una actividad dada la hora de término en un reloj de manecillas.
67	Resolver problemas en los que se calcula la hora de inicio de una actividad, dada la hora de termino y la duración.
68	Resolver problemas en los que se use la regla graduada para medir objetos en centímetros.

La presentación de los valores en la base datos para estos logros académicos se estableció de la siguiente manera.

Tabla 15 Valores posibles en los diferentes logros académicos

valor	Representación
0	Fallo
1	Acierto
5	Reactivo no alcanzado (se considera como no administrado)
6	Reactivo no administrado por error de impresión
7	Reactivo no administrado (No se incluyó en su cuadernillo de examen)

Como lo indica la información presentada por el mismo INNE solo se consideran reactivos administrados, aquellos que tienen fallo o acierto, por lo tanto, a partir de estos reactivos se debe obtener el número total de reactivos administrados para obtener una calificación para cada uno de los individuos. Para calcular la calificación se utilizaron las siguientes fórmulas:

Total de reactivos administrados = reactivos con acierto + reactivo con fallo

$$\text{Calificación} = \frac{\text{reactivos con acierto}}{\text{Total de reactivos administrados}} * 10$$

Con la formula anterior, se obtuvo la calificación correspondiente a cada uno de los alumnos que aplicaron la prueba EXCALE en un rango de 0 a 10 con decimales. A continuación, se procedió a clasificar a los alumnos de acuerdo a su calificación obtenida, para ello, se determinaron las clases de alumno con bueno, regular y mal desempeño, de la siguiente manera:

Tabla 16 Clasificación de alumnos en base a su desempeño

Rango	Desempeño	Identificador
Calificación <= 3.33	Malo	1
3.33 < calificación <= 6.66	Regular	2
6.66 < calificación	Bueno	3

3.3.3. Datos de entrada para el algoritmo

Combinando el resultado de la agrupación de variables para calcular el valor representativo de los constructos y la clasificación de los alumnos en base a su desempeño educativo, obtendremos una tabla, que fue el insumo de entrada para nuestro algoritmo de testores típicos.

En la tabla 17 podemos observar un ejemplo de esa tabla.

Tabla 17 Ejemplo de datos de entrada para el algoritmo de testores

edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	Clase (desempeño)
1	11	1.43	7.25	2	1.2	1	10.83	5	4	1
2	12.11	2.5	7.25	1.75	2	1	10.83	7.5	4.33	2
2	10.17	3	5	0.5	1.2	1	11.25	5.25	4.5	3
1	13.33	2.6	8	2	1.2	0.5	10.92	6.75	5	1
1	12.32	1.68	7	2	0.7	0.5	11	5	5	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nC1	nC2	nC3	nC4	nC5	nC6	nC7	nC8	nC9	nC10	Cn

3.3.4. Equipos y herramientas

- Para el desarrollo de la herramienta y para llevar a cabo el preprocesamiento, un equipo de cómputo con características.
 - Procesador: Intel Core i7 de doble núcleo y 2.3 GHz.
 - RAM: 8.00 gigabyte (GB).
 - Software:
 - SO Windows 7
 - IDE NetBeans para java
- Debido a que se requieren una mayor cantidad de recursos para llevar a cabo la ejecución de la herramienta con los datos preparados, se opta por utilizar un equipo diferente y con mayores capacidades.
 - Procesador: Intel Xeon 4.54 GHz.
 - RAM: 16.00 gigabytes (GB).

- Software:
 - SO Windows Server 2008.
 - IDE NetBeans para correr la herramienta.

3.4. Implementación del algoritmo

Para la realización del algoritmo, se realizó un prototipo, que consistió en crear una solución para una instancia de prueba con pocos registros, con el fin de identificar los potenciales errores o situaciones que se pudiesen presentar con el uso de nuestros datos. En la siguiente figura aparece una secuencia de los pasos que sigue dicha implementación y entre paréntesis se detalla la salida de cada una de las etapas:

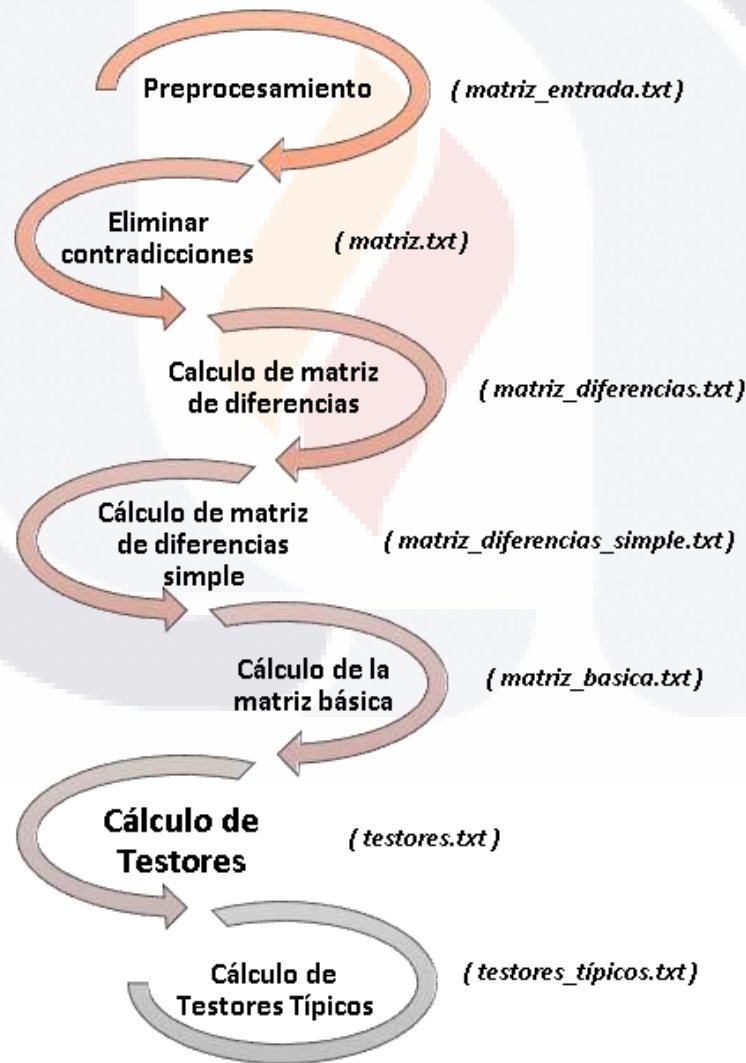


Figura 6 Etapas del algoritmo de Testores.

3.4.1. Elementos del algoritmo desarrollado

En el algoritmo utilizado para el desarrollo de nuestro prototipo, se encuentran los siguientes elementos:

Matriz de entrada (MA)

Esta matriz fue cargada en un archivo de texto plano, y cuenta con un campo por cada variable, además de que contiene un índice compuesto del identificador de la clase.

Matriz de aprendizaje (MA)

Esta matriz es generada en un archivo de texto, y así como la matriz de entrada cuenta con un campo por cada variable, y el identificador de la clase, sin embargo, esta matriz excluye aquellos registros de la matriz de entrada que cuenten con características idénticas, pero que se encuentren en diferente desempeño académico (clase), esto debido a que son contradicciones que afectan el correcto funcionamiento del algoritmo.

Matriz de diferencias (MD)

Esta matriz se generó creando una comparación de todos los elementos de una clase contra todas las demás, la combinación de todas las variables nos dio como resultado la matriz de diferencia, en donde se obtiene un 1 cuando existe diferencia y un 0 cuando no existe.

Matriz de diferencias simple (MDS)

Esta matriz es una réplica de la matriz de diferencias anterior solo que excluyendo aquellos registros que fuesen idénticos en todos sus elementos.

Matriz básica (MB)

Para la generación de esta matriz, se realizó una reducción de la matriz de diferencias simple anterior, en la cual se seleccionaron sólo los registros que son esenciales, eliminando los subconjuntos redundantes.

Matriz de testores (MT)

Como lo dice la literatura, un testor es conjunto de características para identificar a un individuo que pertenece a una clase o a otra, por lo que, en esta parte, se realizó una búsqueda exhaustiva utilizando cada posible variable contenida en un conjunto de 10 bits (uno por cada variable) y se analizó si cumple o no con la condición de ser testor.

Matriz de testores típicos (MTT)

En esta matriz, se realizó una reducción de la matriz de testores anterior, eliminando los subconjuntos, y dejando sólo los sets de datos esenciales, que son los testores típicos.

Para el desarrollo del prototipo se utilizó el lenguaje de programación JAVA, este prototipo recibía como entradas la ruta de la matriz de entrada, el número de características que se toman en cuenta en el archivo, así como el total de registros incluidos en ese archivo. Durante cada fase de la ejecución del prototipo se generaron los siguientes resultados:

matriz.txt: es un archivo de texto plano que contiene la base de datos de la cual previamente se transformaron las variables de trabajo, excluyendo aquellos registros con iguales características, pero ubicados en una clase de desempeño diferente.

matriz_direferencias.txt: archivo que contiene todas las diferencias encontradas entre clases de la MA (Matriz de Aprendizaje) en este caso: *matriz.txt*.

matriz_diferencias_simple.txt: contiene los registros de la matriz de diferencias simple excluyendo los que son repetidos.

matriz_basica.txt: contiene la reducción de los registros en el archivo *matriz_diferencias_simple.txt*, tomando sólo aquellos elementos indispensables.

testores.txt: en esta tabla se encuentran todas las combinaciones posibles de testores.

testores_tipicos.txt: aquí se almacenan los testores típicos.

El primer paso para la resolución consistió en obtener la matriz de aprendizaje en la cual se tienen todas las características (constructos) de los objetos (alumnos) y un índice que consiste en la clase a la que pertenecen. Este resultado se obtuvo de la transformación de

variables de trabajo, así que solo resta exportar este resultado a un archivo de texto nombrándolo *matriz_entrada.txt* separado por comas en una ubicación conocida.

Posterior a esto, fue necesario crear el archivo *matriz.txt* que excluye a aquellos elementos que tienen iguales características, pero están ubicados en diferente clase, esto se realizó mediante los siguientes pasos.

Algoritmo 1 Eliminación de contradicciones

1. Mediante una iteración común, almacenar el contenido de **matriz_entrada.txt** en un arreglo de manera que cada posición contenga un renglón (registro con características y clase) del archivo.
2. Inicializar un índice *i* para recorrer en un bucle la matriz de datos.
 1. Dentro del bucle inicializar un índice *j* para recorrer en un segundo bucle desde la posición *i* de la matriz de datos.
 1. Comparar los valores de la posición *i* y *j* de la matriz de datos, en caso de que alguno de los valores sea nulo, incrementar el índice *i* se rompe el segundo bucle ya que todas las comparaciones siguientes serán siempre innecesarias.
 2. En caso de que ambos valores sean diferentes a nulo, se debe hacer la comparación entre el atributo clase (la última columna en el registro) de estos valores, si ambos contienen la misma clase entonces se incrementa el índice *j* y se continua con las comparaciones.
 3. Después de cerciorarse de que los valores en los índices *i* & *j* son diferentes de nulo y de que los valores en estas posiciones pertenecen a diferentes clases, se procede a comparar las características de cada registro, y si al obtener el registro de diferencias se obtiene un registro que contiene solo ceros (0,0,0,...,0), se asignan valores nulos a las posiciones *i* & *j* para indicar que son posiciones o registros que incluyen contradicciones en su contenido, además se incrementa el índice *i* y se rompe el segundo bucle, ya que las siguientes comparaciones serán innecesarias.
 4. Al saber que el resultado de la comparación entre los registros de las posiciones *i* & *j* no arrojaron un resultado conformado por solo ceros se incrementa el índice *j*
 5. Si el índice *j* es igual al tamaño del arreglo, se rompe el segundo bucle, y se incrementa *i*.
 2. Si el índice *i* es igual al tamaño del arreglo con los valores del archivo inicial, se rompe el bucle.
3. Por último, se procede a almacenar el contenido del arreglo que contiene los valores iniciales en un nuevo archivo de texto (*matriz.txt*), esto excluyendo aquellas posiciones que en su respectivo valor incluyen nulo.

Posterior a estos pasos se obtuvo la matriz de diferencias la cual, almacena en el archivo *matriz_diferencias.txt*, para llenar este archivo se siguieron los siguientes pasos.

Algoritmo 2 Cálculo de matriz de diferencias

1. Mediante una iteración común se almacena el contenido de **matriz.txt** en un arreglo de manera que en cada posición se contenga un renglón (registro con características y clase) del archivo.
2. Se crea el archivo *matriz_diferencias.txt* en la ruta deseada,
3. Inicializar un índice *i* para recorrer en un bucle la matriz de datos.
 1. Dentro del bucle inicializar un índice *j* para recorrer en un segundo bucle desde la posición *i* la matriz de datos.
 1. Comparar los valores de la posición *i* y *j* de la matriz de datos, en caso de que alguno de los valores sea nulo, se incrementa el índice *i* se rompe el segundo bucle ya que todas las comparaciones siguientes serán siempre innecesarias.
 2. En caso de que ambos valores sean diferentes a nulo, se debe hacer la comparación entre el atributo clase (la última columna en el registro) de estos valores, si ambos contienen la misma clase entonces se incrementa el índice *j* y se continua con las comparaciones.
 3. Después de cerciorar que los valores en los índices *i* & *j* son diferentes de nulo y de que los valores en estas posiciones pertenecen a diferentes clases se procede a comparar las características de cada registro, y se almacena el resultado de esta comparación en el archivo *matriz_diferencias.txt*.
 4. Si El índice *j* es igual al tamaño del arreglo, se rompe el segundo bucle, y se incrementa el índice *i*.
 2. Si el índice *i* es igual al tamaño del arreglo con los valores del archivo inicial, se rompe el bucle

Lo siguiente es obtener el archivo *matriz_diferencias_simple.txt*, el cual contendría los valores del archivo *matriz_diferencias.txt* excluyendo aquellos registros que son idénticos.

Algoritmo 3 Cálculo de matriz de diferencias simples

1. Mediante una iteración común se almacena el contenido de **matriz_diferencias.txt** en un arreglo que en cada posición contenga un renglón (registro con diferencias) del archivo.
2. Crear una lista vacía que contenga los valores singulares de la matriz de diferencias.
3. Inicializar un índice *i* para recorrer en un bucle la matriz de diferencias.
 1. A continuación, se compara el registro *i* de la matriz de diferencias con cada una de las posiciones de la lista de diferencias singular y en el caso de no encontrarse en esta lista, se añade, de lo contrario, se omite y solo se incrementa el índice *i*.
 2. Lo anterior repetirlo para los valores de todas las posiciones de la matriz de diferencias.
4. Por último, se procede a almacenar el contenido de la lista que contiene los valores de la matriz de diferencias simple, pero sin repetir en un nuevo archivo de texto (**matriz_diferencias_simple.txt**).

Obtenida la matriz de diferencias sin repetición se procedió a la obtención de la MB, para la cual se realizaron los siguientes pasos:

Algoritmo 4 Cálculo de matriz Básica

1. Mediante una iteración común se almacena el contenido de **matriz_diferencia_simple.txt** en un arreglo que en cada posición contuviese un renglón (registro con diferencias) del archivo.
2. Después de la posterior lectura de la matriz de diferencias, es necesario convertir cada uno de esos registros a su valor entero y almacenarlos en un nuevo vector, esto para facilitar el trabajo, tomando que cada registro es una sucesión de bits separados por comas, de tal manera que representan en binario, el valor de un número en decimal.
3. Para lograr calcular la matriz básica, es necesario tener ordenada la matriz de diferencias, para ello, se puede utilizar cualquier método de ordenamiento, en este caso se utilizó el método de la burbuja.
4. Teniendo el arreglo de números enteros representativos de la matriz de diferencias de forma ordenada y que excluye a las diferencias repetidas se procede a generar la matriz básica, recordando que la matriz básica es la reducción de los registros en el archivo **matriz_diferencias_simple.txt**, tomando sólo aquellos elementos que son indispensables.
5. Para generar la matriz básica es necesario crear un bucle que recorra todo el arreglo de la matriz de diferencias, para ello se crea un índice *i*.
 1. Dentro del bucle inicializar un índice *j* para recorrer en un segundo bucle desde la posición *i+1* la matriz de diferencias.
 1. Es necesario asegurarse de que el valor de la posición *i* en el arreglo de la matriz de diferencias es diferente de nulo, en este caso es necesario incrementar el índice *i* y terminar el segundo bucle.

2. Sabiendo entonces que la posición i del arreglo de la matriz de diferencias es diferente de nulo, es necesario hacer lo mismo para la posición j del arreglo, si es nulo, se incrementa j y se continua a la siguiente iteración.
2. Al saber que los valores en las posiciones i & j son diferentes de nulo se procede a comparar si el valor de la posición j es un subconjunto del valor en la posición i , para esto basta con hacer una operación “**and**” entre los dos valores y si el resultado de esta operación es el mismo valor que contiene el índice i , quiere decir que efectivamente la posición j es un subconjunto de la posición i .
 1. En caso de que el valor de la posición j en el arreglo sea un subconjunto del valor en la posición i , se procede a asignar un valor nulo a la posición j , se incrementa j y se continua con las iteraciones del segundo bucle.
 2. Si El índice j es igual al tamaño del arreglo de la matriz de diferencias, se rompe el segundo bucle, y se incrementa i .
3. Si El índice i es igual al tamaño del arreglo de la matriz de diferencias, se terminan las iteraciones del primer bucle.
6. Por último, se procedía a almacenar el contenido del arreglo con el que se estuvo interactuando en un nuevo archivo de texto (**matriz_basica.txt**), este almacenamiento se debe hacer convirtiendo cada uno de los valores en el vector a su respectivo valor en binario y separando cada bit por una coma, lo anterior excluyendo aquellas posiciones que en su respectivo valor contienen nulo.

Una vez obtenido el archivo con la matriz básica se procedió a obtener los testores a partir de esta matriz, para ello se realizaron los siguientes pasos:

Algoritmo 5 Cálculo de Testores

1. Mediante una iteración común se almacena el contenido del archivo `matriz_basica.txt` en un arreglo, de tal manera que cada posición contuviese un renglón (registro de la matriz básica) del archivo.
2. Después de la posterior lectura de la matriz básica, es necesario convertir cada uno de esos registros a su valor entero y almacenarlos en un nuevo vector, esto para facilitar el trabajo, es necesario tener en cuenta que cada registro es una sucesión de bits separados por comas, de tal manera que representan en binario, el valor de un numero en decimal.
3. Se debe crear una lista vacía para almacenar los testores encontrados.
4. Teniendo el arreglo de números enteros representativos de la matriz básica se procede a la generación de los testores, hay que recordar que un testor es conjunto de características para identificar a un individuo en una clase y es necesario analizar cada variable posible en el conjunto determinado por el total de bits (variables) disponible y determinar una a una si es que se trata de un testor.
 1. Se calcula el número máximo posible a partir de la cantidad de bits que se tienen disponible, basta con realizar la ecuación $2^n - 1$, siendo n el número de bits.
 2. Se inicializa un contador x en 1 y se itera un bucle mientras que x sea menor al número máximo posible.

1. A continuación, se debe evaluar si el valor actual de x cumple con las características de un testor. Para saber si el valor actual de x es o no un testor, es necesario evaluarlo con todos y cada uno de los elementos de la matriz básica.
 1. Es posible decir que x es un testor si al realizar una operación **and** con todos y cada uno de los elementos de la matriz básica, todos los resultados de estas operaciones son diferentes a cero.
 2. En el caso de que x sea un testor, es necesario añadirlo a la lista de testores creada previamente
 3. En caso de que x no sea un testor se incrementa su valor y se vuelve a evaluar este nuevo valor con todos los elementos de la matriz básica.
5. Por último, se procede a almacenar el contenido de la lista que incluye los valores aceptados como testores en un nuevo archivo de texto (`testores.txt`), al igual que en los pasos anteriores este almacenamiento debe hacerse convirtiendo cada uno de los valores en la lista a su respectivo valor en binario y separando cada bit por una coma.

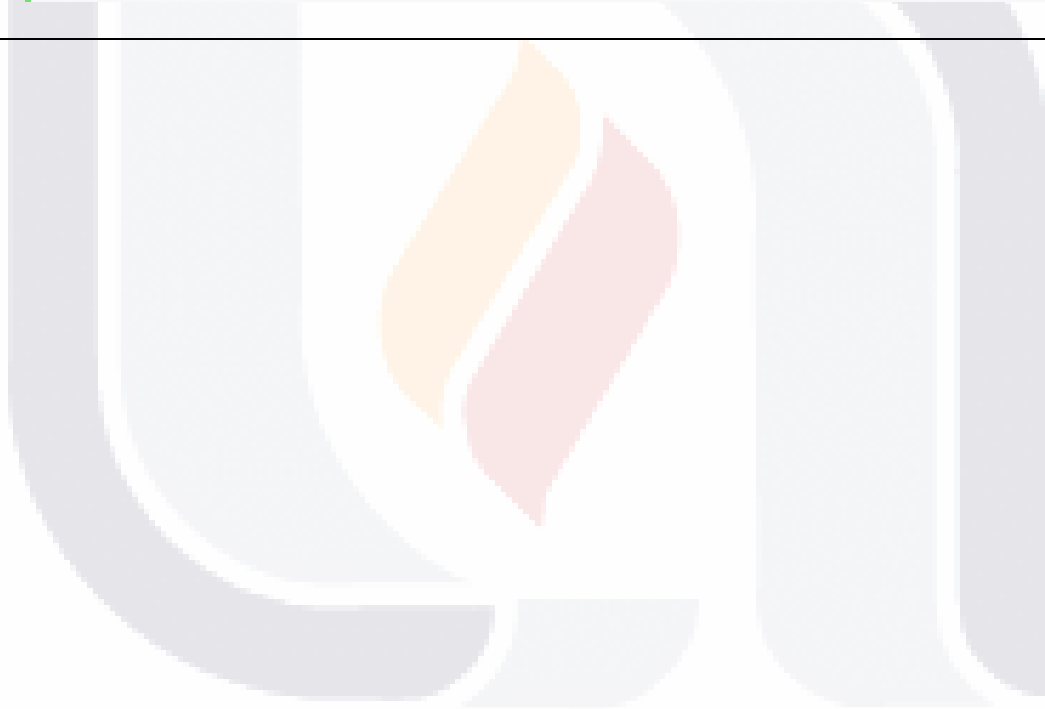
Una vez que obtenidos los testores, y teniéndolos almacenados en el archivo `testores.txt` se debe proceder a la obtención de los testores típicos, para lo cual se realizaron los siguientes pasos:

Algoritmo 6 Cálculo de Testores Típicos

1. Mediante una iteración común se almacenó el contenido de **testores.txt** en un arreglo que en cada posición contuviese un renglón (testor) del archivo.
2. Después de la posterior lectura de los testores, es necesario convertir cada uno de esos registros a su valor entero y almacenarlos en un nuevo vector, esto para facilitar el trabajo, tomando en cuenta que cada registro es una sucesión de bits separados por comas, de tal manera que representan en binario, el valor de un numero en decimal.
3. Teniendo el arreglo de números enteros representativos de los testores obtenidos previamente, se procede a la obtención de los testores típicos, hay que recordar que los testores típicos son aquellos testores que son indispensables y no son subconjuntos de otros.
4. Para obtener los testores típicos es necesario crear un bucle que recorra todo el arreglo de testores, para ello se crea un índice i inicializándolo en 0.
 1. Dentro del bucle, se debe inicializar un índice j para recorrer en un segundo bucle desde la posición $i+1$ el arreglo de testores.
 1. Es necesario asegurarse de que el valor de la posición i en el arreglo de testores es diferente de nulo, en este caso es necesario incrementar el índice i y terminar el segundo bucle.
 2. Sabiendo entonces que la posición i del arreglo de testores es diferente de nulo, se debe hacer lo mismo para la posición j del arreglo, si es nulo, se incrementa j y se continua a la siguiente iteración del siguiente bucle.
 2. Al cerciorarse de que los valores en las posiciones i & j son diferentes de nulo, se procede a comparar si el valor de la posición j es un subconjunto del valor en

la posición i , para esto basta con hacer una operación “*and*” entre los dos valores y si el resultado de esta operación es el mismo valor que contiene el índice i , quiere decir que efectivamente la posición j es un subconjunto de la posición i .

1. En caso de que el valor de la posición j en el arreglo sea un subconjunto del valor en la posición i , se asigna un valor nulo a la posición j , se incrementa j y se continua con las iteraciones del segundo bucle.
2. Si el índice j es igual al tamaño del arreglo de testores, se rompe el segundo bucle, y se incrementa el índice i .
3. Si el índice i es igual al tamaño del arreglo de testores, se terminan las iteraciones del primer bucle.
5. Por último, se procede a almacenar el contenido del arreglo con el que se estuvo interactuando en un nuevo archivo de texto (testores_tícos.txt), este almacenamiento se debe hacer convirtiendo cada uno de los valores en el vector a su respectivo valor en binario y separando cada bit por una coma, lo anterior excluyendo aquellas posiciones que en su respectivo valor contienen nulo.



3.5. Construcción de la herramienta para validación

Una vez obtenidos los testores típicos, evaluamos su eficiencia, para ello, fue necesario crear una herramienta capaz de clasificar los datos en base a las características de los testores encontrados en cada una de las ejecuciones del algoritmo, de tal manera que para cada uno de los testores sea posible, crear un modelo diferente de clasificación, esto debido a que individualmente cada testor representa un subconjunto diferente de características de la base de datos principal.

En la siguiente imagen se muestra la herramienta creada en KNIME para llevar a cabo dicha validación:

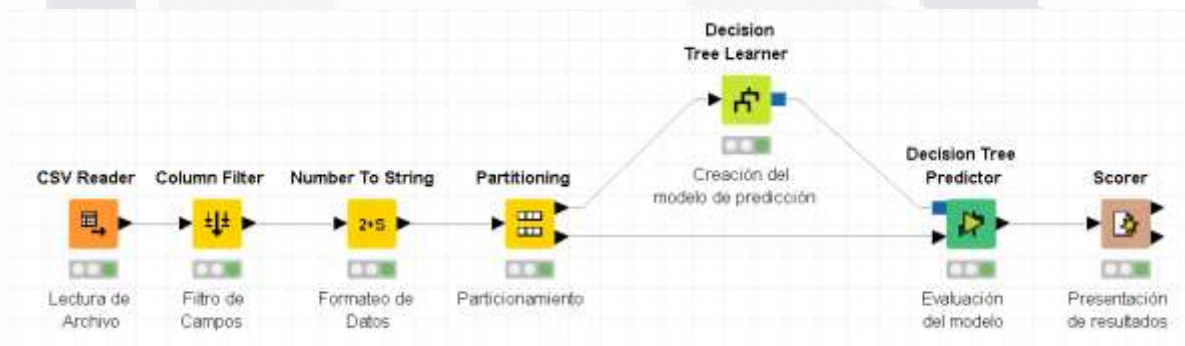


Figura 7 Herramienta de validación creada en KNIME.

Las diferentes etapas de esta herramienta se explican a continuación:

Lectura del archivo

En esta etapa con ayuda del componente “CSV Reader” se lleva a cabo la lectura del archivo de datos, el cual contiene el total de todos nuestros datos, mismos que se utilizaron para leer obtener los testores típicos.

Filtro de Campos

Debido a que cada testor típico representa un grupo diferente de características, es necesario diferenciar aquellas que pertenecen al testor típico en cuestión y partir de ello, generar un nuevo grupo de datos el cual representará al testor con el cual se esté trabajando.

Formateo de datos

El archivo de datos de entrada incluye dentro de su contenido tanto a los constructos o características como a las clases a las que pertenecen cada uno de los registros, sin embargo, todo el contenido está representado de forma numérica, es por esto que es necesario convertir tanto la clase como algunos de los campos a su valor *String*, esto para que se conviertan en tipos de dato Nominal, ayudando así a facilitar la creación del modelo de clasificación,

Particionamiento

Es necesario separar aleatoriamente el conjunto de datos suministrado en dos subconjuntos diferentes, de los cuales uno contendrá el 80% de los datos para llevar a cabo el entrenamiento del modelo de clasificación, mientras que el otro subconjunto de datos contendrá el 20% de los datos que serán de utilidad para evaluar el modelo creado con el primer conjunto.

Creación del modelo de Predicción

Con la ayuda del componente “Decision Tree Learner” en KNIME, es posible generar el un modelo para la clasificación del 80 % de los datos particionados anteriormente, este conjunto de datos representa la entrada de entrenamiento para que el modelo de predicción logre aprender a clasificar. En este componente es necesario indicar cuál de los campos proporcionados representa a nuestra clase en la cual se clasifican cada uno de nuestros registros.

Evaluación del Modelo

Una vez creado nuestro árbol o modelo de clasificación con el 80 % de los datos introducidos, se utilizará el 20 % restante para evaluar la fiabilidad de este modelo, con ayuda del componente “Decision Tree Predictor” se tratará de clasificar los registros de este nuevo subconjunto de datos en base a las características que lo componen, una vez clasificados en base al modelo, se debe hacer una comparación con la clase a la cual pertenece el registro y la clase en la cual fue clasificado por el modelo.

Presentación de resultados

Para llevar a cabo la comparación entre la clase en la cual el modelo clasificó a los datos del segundo subconjunto de datos que contienen el 20% del conjunto original, y la clase a la que realmente pertenecen, es necesario llevar a cabo una contabilización de esos datos y para ello el componente “scorer” de KNIME, el cual presenta múltiples herramientas que permiten visualizar, comparar y evaluar diferentes conjuntos de datos y a partir de ello arrojar información que sea de utilidad.



4. RESULTADOS

4.1. Identificación de factores para la matriz de entrada

El formato de los datos de entrada para la ejecución del algoritmo a nuestra matriz de entrada se observa en la siguiente tabla.

Tabla 18 Datos de entrada para la ejecución del algoritmo.

edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	Clase (desempeño)
1	11	1.43	7.25	2	1.2	1	10.83	5	4	1
2	12.11	2.5	7.25	1.75	2	1	10.83	7.5	4.33	2
2	10.17	3	5	0.5	1.2	1	11.25	5.25	4.5	3
1	13.33	2.6	8	2	1.2	0.5	10.92	6.75	5	1
1	12.32	1.68	7	2	0.7	0.5	11	5	5	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nC2	nC3	nC4	nC5	nC6	nC7	nC8	nC9	nC10	nC11	Cn

Al aplicar el algoritmo a nuestra matriz de entrada, se obtuvo un solo testor y por ende un solo testor típico. El testor obtenido fue el número 1023, que en binario puede escribirse de la siguiente manera:

Tabla 19 Testores obtenidos en la ejecución del algoritmo.

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos
1023	1	1	1	1	1	1	1	1	1	1

Lo que este testor representa es que las 10 características en su conjunto son necesarias e indispensables para lograr clasificar y/o predecir el desempeño educativo de los alumnos en la prueba Excale.

4.1.1. Análisis de resultados para la matriz de entrada

Tomando en cuenta que todas las variables son necesarias e indispensables para lograr clasificar y predecir el desempeño de los alumnos en la prueba EXCALE, debemos incluir todas estas variables en nuestro modelo en KNIME.

Después de realizar la evaluación en el modelo de KNIME, se encontró que sólo fue posible clasificar un 33.8% de los datos en el subconjunto reservado para la evaluación del modelo, indicando así que no es muy eficaz la clasificación de los alumnos en base a su desempeño educativo, al menos de esta manera.

4.2. Identificación de factores para la matriz de entrada incluyendo el campo entidad

Para lograr obtener resultados que sean de mayor utilidad se realizó una segunda ejecución en el proyecto, pero esta vez incluyendo una nueva variable nuestro conjunto de datos; para esta ejecución se agregó el campo entidad, un dato expresado de forma numérica, que tanto la evaluación de matemáticas en el EXCALE 03_2014 como la evaluación de contexto incluyen. En la siguiente tabla se muestra el formato de los nuevos datos de entrada para el algoritmo.

Tabla 20 Datos de entrada para la ejecución del algoritmo (incluyendo al campo entidad)

entidad	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	Clase (desempeño en la prueba)
1	1	11	1.43	7.25	2	1.2	1	10.83	5	4	1
23	2	12.11	2.5	7.25	1.75	2	1	10.83	7.5	4.33	2
56	2	10.17	3	5	0.5	1.2	1	11.25	5.25	4.5	3
23	1	13.33	2.6	8	2	1.2	0.5	10.92	6.75	5	1
2	1	12.32	1.68	7	2	0.7	0.5	11	5	5	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nC1	nC2	nC3	nC4	nC5	nC6	nC7	nC8	nC9	nC10	nC11	Cn

La siguiente tabla muestra el resultado obtenido al ejecutar el algoritmo incluyendo la variable entidad en el conjunto de datos.

Tabla 21 Testores obtenidos en la ejecución del algoritmo (incluyendo al campo entidad).

Testor	entidad	alumno	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	Clase (desempeño)
2047	1	1	1	1	1	1	1	1	1	1	1	1	1

4.2.1. Análisis de resultados para la matriz de entrada incluyendo la variable entidad

Al igual que para la primera ejecución, en esta ocasión, el resultado fue un testor típico, representado por la secuencia de bits encendidos(unos) que indican que todas y cada una de las variables son necesarias para lograr definir en el desempeño educativo de los alumnos en la prueba EXCALE.

Después de realizar la prueba de evaluación en el modelo en KNIME con todas y cada una de las variables, incluyendo al campo entidad, se encontró que solo fue posible clasificar un 44.4% del subconjunto de datos reservado para evaluación, esto indica que el modelo utilizado no es lo suficientemente eficaz para clasificar a los alumnos en base a su desempeño educativo, inclusive, incluyendo el campo entidad en el conjunto de datos.

4.3. Identificación de factores que determinan el desempeño de un alumno en la prueba EXCALE en el estado de Aguascalientes

4.3.1. Identificación de testores típicos en el estado de Aguascalientes

Para obtener información más específica, se tomó la decisión de hacer una tercera ejecución que contuviese solo los registros de una entidad, para esta ejecución se eligieron los registros correspondientes al estado de Aguascalientes, es decir seleccionar solo aquellos resultados donde su entidad fuese 1 que representa a Aguascalientes en la documentación de los datos.

Después de llevar a cabo esa ejecución se obtuvieron 94 testores de los cuales 12 eran testores típicos, en la siguiente tabla se observan los 12 testores típicos con su expresión en binario representando cada uno a diferentes características o constructos.

Tabla 22 Testores obtenidos en la ejecución correspondiente al estado de Aguascalientes.

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos
255		1	1	1	1	1	1	1	1	1
343		1		1		1		1	1	1
374		1				1			1	
391		1	1					1	1	1
406		1	1			1		1	1	
413		1	1			1	1	1		1
429		1	1		1		1	1		1
469		1	1	1		1		1		1
483		1	1	1	1				1	1
485		1	1	1	1			1		1
823	1	1			1	1		1	1	1
853	1	1		1		1		1		1

4.3.2. Análisis de resultados para el estado de Aguascalientes

Tomando en cuenta que cada testor típico representa una alternativa diferente para clasificar nuestra información, se debe realizar una corrida diferente de nuestra herramienta creada en KNIME para validación. Al realizar las diferentes corridas, se obtuvieron los siguientes resultados donde el porcentaje de aciertos es el porcentaje de los datos que fue posible clasificar correctamente, y por el otro lado el porcentaje de error es el porcentaje de datos que no fueron posibles de clasificar.

Tabla 23 Porcentaje de clasificación en KNIME en el estado de Aguascalientes.

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	% Acierto	% Error
343		1		1		1		1	1	1	78.4%	21.6%
274		1				1			1		81.1%	18.9%
391		1	1					1	1	1	75.6%	24.4%
406		1	1			1		1	1		70.2%	29.8%
413		1	1			1	1	1		1	70.2%	29.8%
429		1	1		1		1	1		1	81.1%	18.9%
469		1	1	1		1		1		1	67.5%	32.5%
483		1	1	1	1				1	1	62.1%	37.9%
485		1	1	1	1			1		1	75.6%	24.4%
823	1	1			1	1		1	1	1	67.5%	32.5%
853	1	1		1		1		1		1	78.4%	21.6%
peso informacional	18%	100%	64%	45%	36%	64%	18%	82%	55%	82%		

Como observamos en la tabla anterior, los testores típicos con mayor porcentaje de acierto son el testor 274 y el 429, además de esto, es muy notorio el peso informacional del 100% en la variable vivienda, esto quiere decir que esta variable, es fundamental para lograr clasificar a los alumnos, incluso es posible decir que es la variable que más afecta en el desempeño en la prueba EXCALE.

Tomando como estudio el testor 274, uno de los testores con mayor porcentaje de clasificación correcto y además, uno de los que cuentan con menos cantidad de variables, la herramienta de clasificación en KNIME proporciona un árbol bayesiano de clasificación

como el de la siguiente figura, en el área sombreada se encuentran aquellas secciones del árbol que aún se desglosa en ramas más específicas.

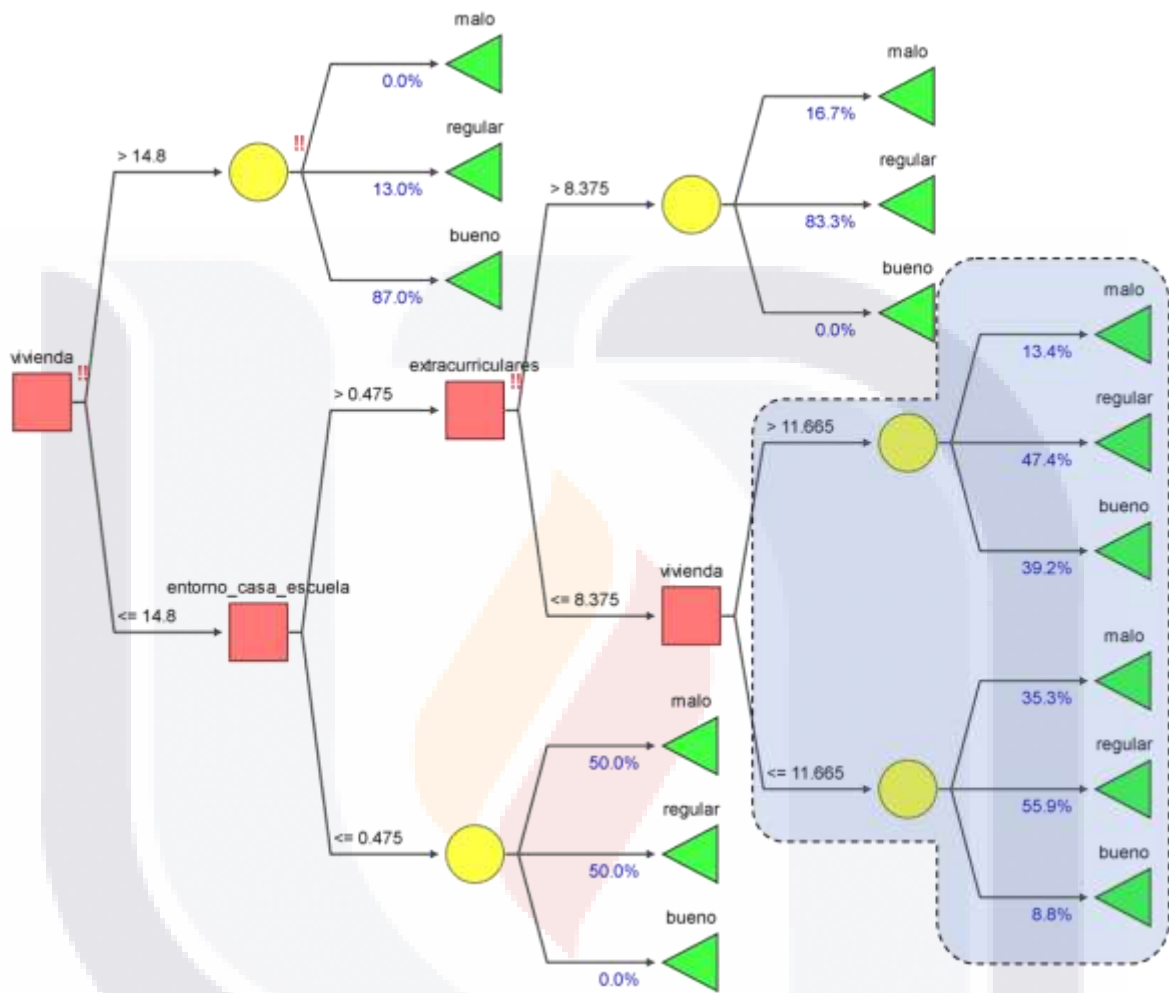


Figura 8 Árbol bayesiano del estado de Aguascalientes.

4.4. Identificación de factores que determinan el desempeño de un alumno en la prueba EXCALE en el estado de Jalisco

4.4.1. Identificación de testores típicos en el estado de Jalisco

Así como la prueba realizada para el estado de Aguascalientes, se decidió realizar este mismo procedimiento para otras entidades, de entre ellas, el estado de Jalisco, así pues, se seleccionaron solo aquellos datos pertenecientes a los alumnos que presentaron la prueba EXCALE en este estado, señalados en la base de datos con un 14 según la documentación.

Después de llevar a cabo esa ejecución se obtuvieron 53 testores de los cuales 7 eran testores típicos, en la siguiente tabla se observan estos testores típicos con su expresión en binario, además de las características que representan cada uno de ellos.

Tabla 24 Testores obtenidos en la ejecución correspondiente al estado de Jalisco

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos
215	0	0	1	1	0	1	0	1	1	1
343	0	1	0	1	0	1	0	1	1	1
350	0	1	0	1	0	1	1	1	1	0
455	0	1	1	1	0	0	0	1	1	1
467	0	1	1	1	0	1	0	0	1	1
470	1	0	1	1	0	1	0	1	1	0
1006	1	1	1	1	1	0	1	1	1	0

4.4.2. Análisis de resultados para el estado de Jalisco

Así como en el estado de Aguascalientes, se realizaron pruebas en KNIME para evaluar la eficacia de clasificación en cada uno de los testores típicos obtenidos en el grupo de datos pertenecientes al estado de Jalisco; el porcentaje de acierto y de error obtenido en KNIME, se puede observar en la siguiente tabla.

Tabla 25 Porcentaje de clasificación en KNIME en el estado de Jalisco.

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	% Acierto	% Error
215	0	0	1	1	0	1	0	1	1	1	70.5%	29.5%
343	0	1	0	1	0	1	0	1	1	1	74.4%	25.6%
350	0	1	0	1	0	1	1	1	1	0	87.1%	12.9%
455	0	1	1	1	0	0	0	1	1	1	68.6%	31.4%
467	0	1	1	1	0	1	0	0	1	1	65.2%	34.8%
470	1	0	1	1	0	1	0	1	1	0	77.3%	22.7%
1006	1	1	1	1	1	0	1	1	1	0	63.8%	36.2%
peso informacional	29%	71%	71%	100%	14%	71%	29%	86%	100%	57%		

A diferencia de los resultados obtenidos en el grupo de datos de Aguascalientes, en este nuevo grupo perteneciente al estado de Jalisco, existen dos constructos con un peso informacional del 100%, es decir que estos dos constructos son indispensables para lograr clasificar a los alumnos en base a su desempeño educativo, así mismo se observó que al menos en un testor típico (350), el modelo es muy eficaz al momento de clasificar.

Así pues, tomando como estudio al testor 350 se obtuvo un árbol bayesiano como el que aparece en la siguiente figura:

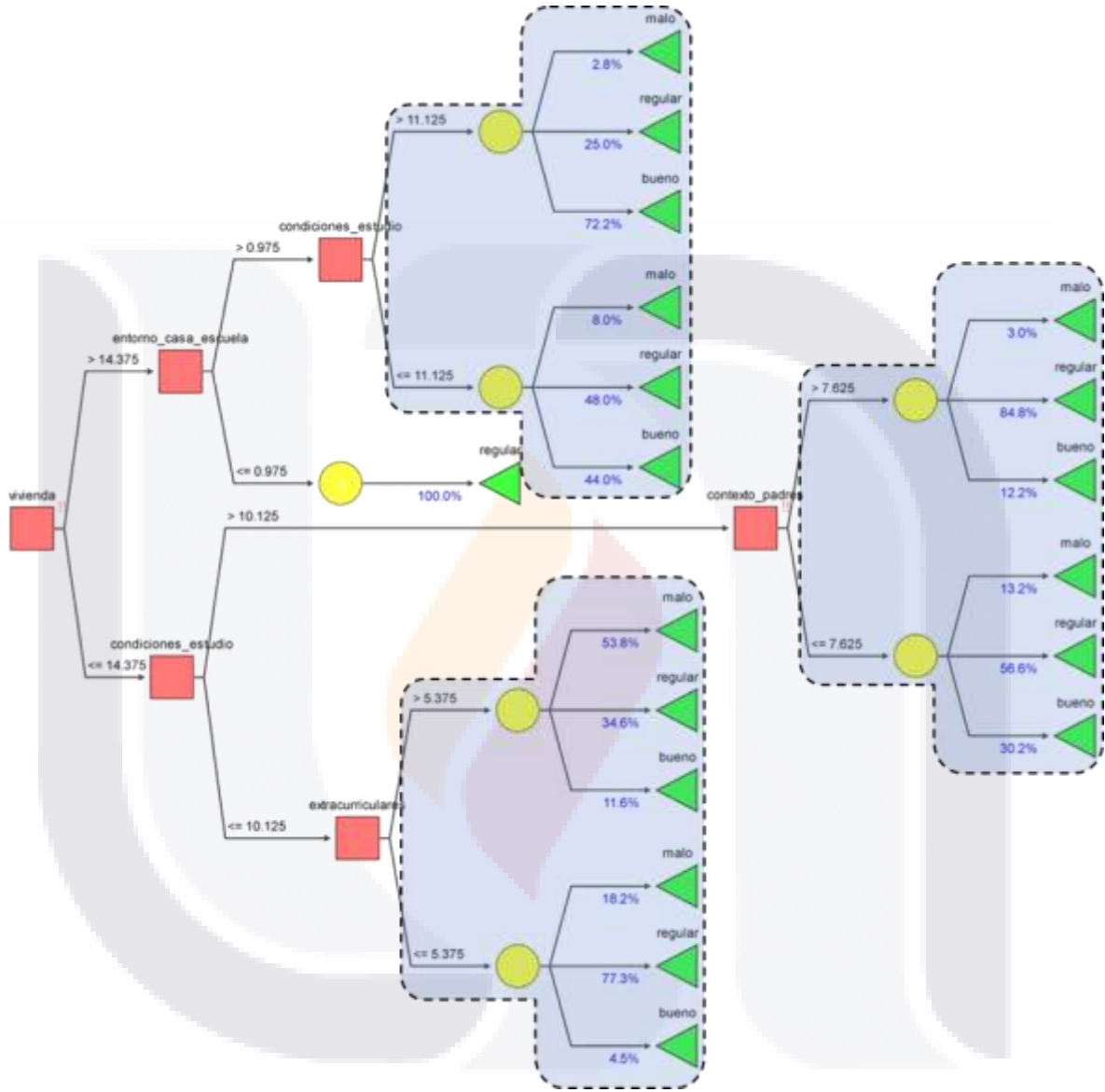


Figura 9 Árbol bayesiano para el estado de Jalisco.

4.5. Identificación de factores que determinan el desempeño de un alumno en la prueba EXCALE en el estado de Veracruz

4.5.1. Identificación de testores típicos en el estado de Veracruz

Conociendo la eficacia en la clasificación de datos en las pruebas del estado de Aguascalientes y de Jalisco, se tomó la decisión de realizar una ejecución más en un estado diferente, para ello se eligió el estado de Veracruz, representado por un 30 en la documentación; esta elección debido a su ubicación geográfica y a la cantidad de sustentantes en la prueba Excale.

Después de llevar a cabo esa ejecución se obtuvieron 54 testores de los cuales 9 eran testores típicos, en la siguiente tabla se observan los 9 testores típicos con su expresión en binario y las características a las cuales representan.

Tabla 26 Testores obtenidos en la ejecución correspondiente al estado de Veracruz.

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos
343	1	0	0	1	0	1	0	1	1	1
359	0	1	0	1	1	0	0	1	1	1
382	0	1	0	1	1	1	1	1	1	0
423	0	1	1	0	1	0	0	1	1	1
455	0	1	1	1	0	0	0	1	1	1
469	0	1	1	1	0	1	0	1	0	1
478	0	1	1	1	0	1	1	1	1	0
508	0	1	1	1	1	1	1	1	0	0
919	1	1	1	0	0	1	0	1	1	1

4.5.2. Análisis de resultados para el estado de Veracruz

De igual manera que en los estados anteriores, se realizó la prueba de validación en KNIME con el grupo de datos del estado de Veracruz, esto para determinar el porcentaje de eficacia de cada uno de los testores típicos, en la siguiente tabla se observan los resultados de esta prueba:

Tabla 27 Porcentaje de clasificación en KNIME en el estado de Veracruz.

Testor	edad_normativa	vivienda	situacion_economica	contexto_padres	futuro_intelectual	entorno_casa_escuela	lengua_indigena	condiciones_estudio	extracurriculares	antecedentes_academicos	% Acierto	% Error
343	1	0	0	1	0	1	0	1	1	1	77.9%	22.1%
359	0	1	0	1	1	0	0	1	1	1	77.5%	22.5%
382	0	1	0	1	1	1	1	1	1	0	73.1%	26.9%
423	0	1	1	0	1	0	0	1	1	1	75.5%	24.5%
455	0	1	1	1	0	0	0	1	1	1	68.6%	31.4%
469	0	1	1	1	0	1	0	1	0	1	73.6%	26.4%
478	0	1	1	1	0	1	1	1	1	0	70.5%	29.5%
508	0	1	1	1	1	1	1	1	0	0	76.3%	23.7%
919	1	1	1	0	0	1	0	1	1	1	71.9%	28.1%
peso informativa	22%	89%	67%	78%	44%	67%	33%	100%	78%	67%		

En esta tercera ejecución, se observa un constructo con un peso informativa del 100%, demostrando que al menos un este constructo es indispensable para llevar a cabo la clasificación de los datos. Además de esto, se puede observar que, aunque el porcentaje de

clasificación correcto es alto en los testores típicos, el modelo no es igual de eficaz que en las dos pruebas anteriores.

Así como en las pruebas anteriores, para el estado de Veracruz se decidió tomar como caso de estudio uno de los testores típicos con mayor porcentaje de clasificación, para este caso se eligió el testor 359, ya que incluye el campo entidad como en los otros dos estudios. El árbol bayesiano con nuestra herramienta en KNIME se muestra en la siguiente figura:

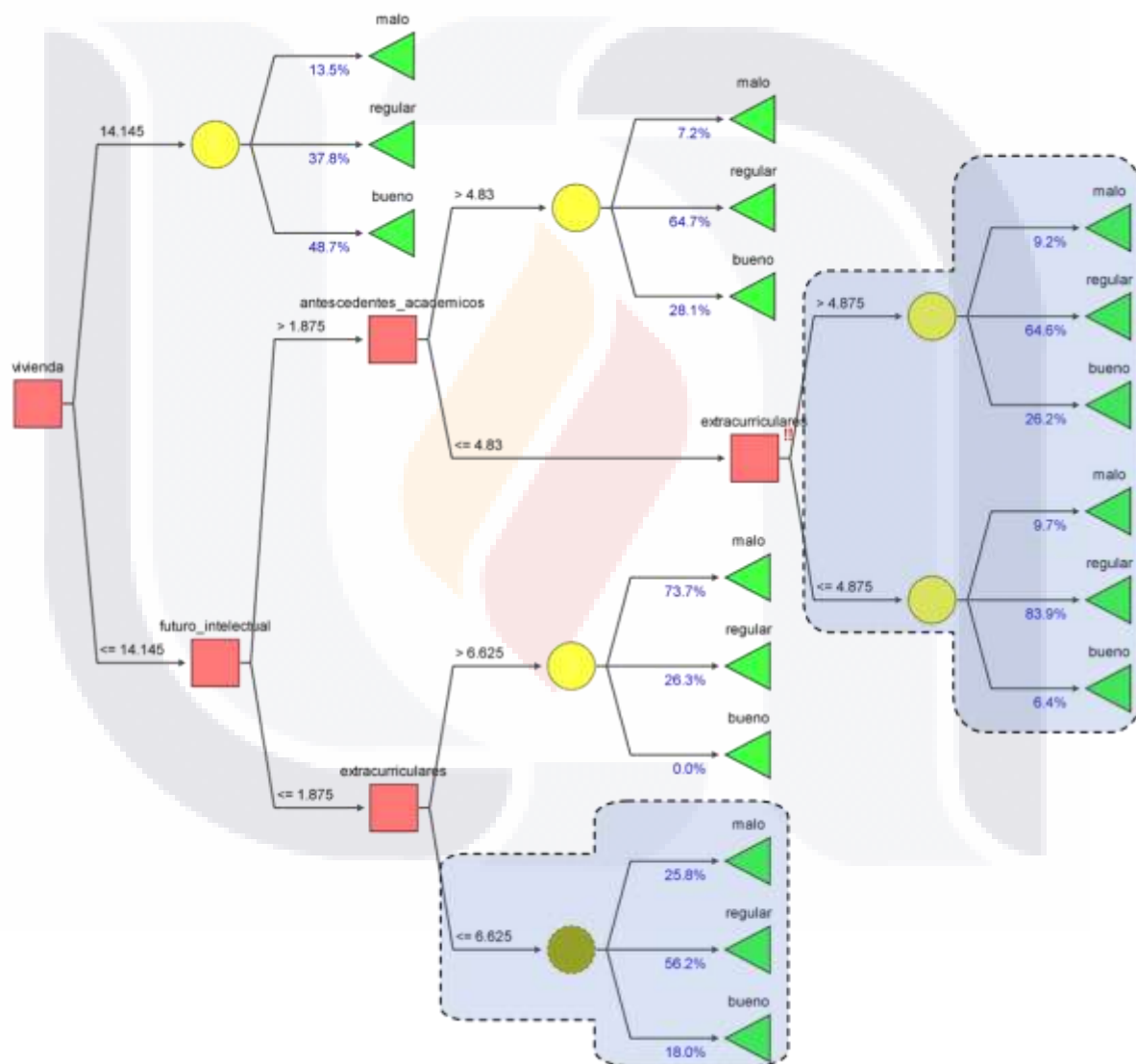


Figura 10 Árbol bayesiano para el estado de Veracruz

5. CONCLUSIONES

Con la realización de este proyecto de tesis, se aprecia el amplio panorama del alcance de las ciencias computacionales debido a su influencia en otras áreas del conocimiento, provocando cambios importantes en sus respectivas investigaciones. Tal es el caso de la educación, la cual, siempre se ha visto beneficiada por la innovación y los avances en ciencia y tecnología que proveen mayores y mejores herramientas que de alguna manera propician una mejora en todas sus vertientes. Por otro lado, estos avances apoyan a la educación en la generación y administración de la información generada.

Gracias a esta interacción, en la actualidad se cuenta con grandes cantidades de información por lo que se requiere de mecanismos más aptos para su análisis. Por esta razón, la inteligencia artificial es el área indicada al proponer nuevas y más eficientes formas de extraer información y generar el conocimiento adecuado para la toma de decisiones. Ejemplo de ello se tiene la selección de características, utilizados en este proyecto, como parte del enfoque lógico-combinatorio del reconocimiento de patrones aplicando la teoría de testores.

La asimilación del funcionamiento y las propiedades de esta teoría de testores, permiten ver los grandes alcances que pudiera tener una rama pequeña de la inteligencia artificial en cualquiera de las otras ramas del conocimiento. Tal fue el caso de este proyecto, en el que se analizaron los resultados de la evaluación de la prueba EXCALE, se localizaron los testores típicos que representan la información mínima para clasificar objetos en sus clases correspondientes.

Por otra parte, no hay que perder de vista que la búsqueda exhaustiva de testores típicos es solo una parte de la investigación, fue necesario determinar cómo interactuaban entre sí las características de los testores para poder así generar clasificadores que ayuden en un futuro a predecir el desempeño educativo de los sustentantes de esta prueba. por lo que se requirió del apoyo de más herramientas de la inteligencia artificial para permitir la creación de estos clasificadores.

La teoría de testores fue la base crear una herramienta que después de validar su correcto funcionamiento, permitió encontrar los testores típicos de cualquier subconjunto de

datos de la prueba EXCALE que se le ingresara, a partir de esto, la herramienta arrojaba solo un testor típico para el conjunto total de toda la prueba, este testor era siempre aquel conformado por todas y cada una de las características, mientras que ingresando subconjuntos delimitados por la entidad a la que pertenecen los sustentantes de la prueba, la herramienta arrojó un rango de testores típicos más amplio.

Para constatar que los testores típicos realmente son conjuntos de características capaces de definir el desempeño educativo de los sustentantes de esta prueba y que además son capaces de predecirlo en pruebas futuras, se hizo uso de una herramienta auxiliar capaz de crear modelos de clasificadores de datos a partir de un conjunto de datos de entrenamiento y un conjunto para evaluar el modelo.

Teniendo entonces los modelos de clasificación y/o predicción, se comprueba el alcance y el beneficio que puede traer la realización de proyectos que involucren la interacción de diferentes áreas como lo es la educación con el análisis de información y la teoría de testores.

5.1. Objetivos cubiertos

En este apartado se describen los apartados de este trabajo de tesis en los que se cubren cada uno de los objetivos definidos en el apartado 1.2.

Objetivo general

Cubierto propiamente por parte del desarrollo y la aplicación de la metodología descrita en el apartado 3. Por su parte, dicha metodología cumple el objetivo en los apartados 3.4 y 3.5, referentes a la implementación del algoritmo, herramienta de validación.

Objetivo Específico 1

Este objetivo se cubre a lo largo del documento, principalmente en el apartado 3 donde se explican las diferentes etapas de la metodología diseñada, así como su función dentro de la investigación.

Objetivo específico 2

La aplicación construida para la búsqueda de los testores típicos que cubrió el objetivo específico número 2, se presenta en el apartado 3.4 que hace referencia a la implementación del algoritmo, donde se detallan todas y cada una de las fases que este sigue para lograr lo prometido. Por su parte, el apartado 3.5 presenta el desarrollo de herramienta en KNIME que mediría la eficacia de esta ampliación a la hora de obtener los testores típicos.

Objetivo Específico 3

Los testores obtenidos después de la implementación de la aplicación desarrollada en el apartado 3.4 se pueden ver representados e interpretados en el apartado 4 de análisis de resultados, donde se explica claramente la representación de cada uno de los testores típicos obtenidos en las diferentes pruebas realizadas con la aplicación, así como el peso informacional de cada una de las variables que componen los conjuntos de datos.

Objetivo específico 4

La herramienta propuesta e implementada para evaluar la eficacia de los testores típicos, así como para explicar el comportamiento de las características que definen el desempeño de los alumnos fue desarrollada en KNIME, dicho desarrollo y su funcionamiento es descrito en el apartado 3.5 y, a su vez, se explican sus resultados en los puntos del apartado 4.

Objetivo específico 5

La presentación y validación de los resultados se presentan en el apartado 4 correspondiente a los resultados, la explicación y estudio de los resultados obtenidos con la herramienta creada en KNIME se presentan en puntos secundarios del mismo apartado 4, en donde se explican el porcentaje de eficacia de cada uno de los testores típicos obtenidos en cada una de las pruebas realizadas, así como los árboles de clasificación obtenidos de algunos testores típicos con mayor porcentaje de eficacia, estos árboles representan rutas de clasificación y o predicción para el desempeño educativo de los sustentantes de la prueba.

5.2. Contribuciones

La principal contribución de este trabajo de tesis es el diseño, implementación y validación de una metodología capaz de identificar el total de los testores típicos en un conjunto de datos de la prueba EXCALE, los cuales representarían los conjuntos mínimos de variables necesarias para definir las diferentes clases que contienen el conjunto de datos seleccionado, así mismo esta metodología sería capaz de identificar el porcentaje de eficacia de cada uno de los testores típicos al momento de intentar definir el comportamiento de los sustentantes en la prueba, además de proveer árboles de decisión bayesianos, los cuales presentarían una explicación de este comportamiento y expondrían un clasificador y/o predictor del desempeño en base a sus características, esto por cada testor típico descubierto.

Esta metodología, es capaz de trabajar para grupos distintos de información limitados por diferentes características, genera al igual diferentes formas de clasificar y/o predecir el desempeño educativo, una forma diferente por cada uno de los testores que se obtengan en su implementación, así pues, cada una de estas formas de clasificación explica la relación de características que comprenden y/o definen una clase específica dentro del conjunto de datos.

Así pues, al implementar esta metodología desarrollada con el conjunto de datos obtenidos en matemáticas de la prueba EXCALE_03_2014 se encontró que utilizando el conjunto de datos completo no es posible determinar un conjunto significativo de testores típicos que representen con eficacia el comportamiento de los datos, sin embargo al delimitar los datos por algunas de sus características como lo es el campo entidad, se encuentran resultados positivos, es decir que se descubren un mayor número de testores típicos, de los cuales, en su mayoría representan un porcentaje muy alto en cuanto a su eficacia de clasificación y predicción. Ejemplo de estas implementaciones con grupos de datos limitados por el campo entidad son las que se presentan en esta investigación, donde se hicieron pruebas y análisis con los resultados de la prueba en las entidades de Aguascalientes, Jalisco y Veracruz.

Para el estado de Aguascalientes, se pueden observar los testores típicos obtenidos, en el apartado 4.3.1 , así mismo la representación y el análisis de estos testores se pueden ver en el apartado 4.3.2, más específico en la figura 8, donde podemos observar que aquellos

alumnos con un alto valor en su característica vivienda es casi imposible ser alumnos de bajo desempeño y además la probabilidad de ser alumnos de alto desempeño es de aproximadamente un 87 %, sin embargo aquellos alumnos que tienen un bajo valor en el constructo casa escuela no tienen muchas probabilidades de ser alumnos de alto desempeño, otra característica especial presentada en la figura 8 es que aquellos alumnos que realizan un alto número de actividades extracurriculares, no necesariamente obtendrán un alto desempeño en la prueba pero posiblemente tampoco serán de bajo desempeño (83% desempeño medio), características como estas son algunas de las contribuciones que deja este trabajo en cuanto al estado de Aguascalientes.

Así como en el estado de Aguascalientes, en el estado de Jalisco, se obtuvieron los testores típicos representativos del conjunto de sustentantes de la prueba EXCALE en esta entidad, y también, se obtuvo un árbol de aquel testor típico que representaría el comportamiento de los sustentantes de la prueba en esta entidad, el cual se presenta en la figura 9, este árbol de clasificación presenta que en conjunto, tener altos valores en los constructos de vivienda, entorno casa escuela y condiciones de estudio pueden indicar una alta probabilidad de ser un estudiante de alto desempeño (72%), y que si al contrario, se tienen pésimos valores en la vivienda y las condiciones de estudio, difícilmente puede ser un alumno de alto desempeño (4.5% y 11.6% dependiendo del valor en las actividades extracurriculares).

Al igual que las otras entidades, el estado de Veracruz arrojó sus propios testores típicos, y por ello su propio árbol bayesiano de clasificación, el cual se presenta en la figura 10, tomando el testor con mayor porcentaje de eficacia podemos observar que aquellos alumnos con bajo valor en los constructos de vivienda, futuro intelectual y un alto valor en las actividades extracurriculares, muy difícilmente será un alumno de alto desempeño, y muy probablemente (73%) será un alumno de bajo desempeño, siendo muy específico, el árbol explica que teniendo un valor bajo en los constructos vivienda y antecedentes académicos y aun teniendo altas expectativas sobre su futuro intelectual, el alumno muy probablemente no será de alto desempeño (6% y 26% dependiendo del valor en las actividades extracurriculares).

Así pues, se encontraron los testores típicos para cada conjunto de datos delimitado por el campo entidad permitió encontrar las características que con mayor eficacia definen el comportamiento de los sustentantes en la prueba EXCALE. En este caso se explicaron solo aquellos testores típicos con mayor porcentaje de eficacia. Para el caso de Aguascalientes se encontraron 12 testores típicos, para el estado de Jalisco 7 y para Veracruz 9. Para cada uno de estos testores típicos se puede definir y explicar un árbol de clasificación que sea capaz de clasificar a los alumnos en un alto, medio y bajo desempeño. Explicado esto, podemos aceptar la hipótesis y aseverar que mediante la lógica combinatoria (teoría de testores), es posible obtener un conjunto de características que distinguen a los estudiantes de alto, medio y bajo desempeño académico en la prueba Excale de Matemáticas en su edición EXCALE 03_2014.

5.3. Trabajo futuro

De acuerdo al análisis de los resultados obtenidos, podemos aseverar sin duda alguna que uso de la teoría de testores puede ayudar a predecir el desempeño de los alumnos en la prueba EXCALE, así pues, se pudiera realizar un estudio de predicción con alumnos próximos a presentar la prueba y con ello predecir su posible desempeño, teniendo así herramientas que permitan mejorar los resultados en la prueba.

Por otra parte con la metodología propuesta, se pudieran realizar estudios, aplicando los mismos procesos pero con conjuntos de datos diferentes, es decir pudieran utilizarse como grupo de estudio, alumnos correspondientes a otras asignaturas en la misma prueba o bien, con alumnos de otras ediciones de la prueba EXCALE, esto pudiera arrojar información significativa en cuanto a las características que definen el desempeño de los alumnos en distintas generaciones, y analizar si es que existen diferencias o similitudes en cuanto a estas características que los definen.

6. BIBLIOGRAFÍA

- Grupo de Tecnologías Avanzadas en Computación. (2008). *Clasificación Supervisada y No Supervisada*. Ecuador: Loxa.
- Lias-Rodríguez, A., & Pons-Porrata, A. (s.f.). *Un nuevo Algoritmo de Escala Exterior para el Cálculo de los Testores Típicos*. Cuba. Obtenido de http://www.cerpamid.co.cu/sitio/files/publicaciones/1034921953BR_RECPAT09.pdf
- Acosta, A. E. (2018). *Identificación de factores de riesgo en patologías médicas mediante métodos de selección de subconjuntos de características*. Aguascalientes : Universidad Autónoma de Aguascalientes.
- Asamblea General de las Naciones Unidas. (1948). Artículo 26. En *Declaración Universal de Derechos Humanos*.
- Asamblea General de las Naciones Unidas. (1959). *declaración de los Derechos del Niño*.
- Backhoff Escudero, E., Bouzas Riaño, A., Hernández, E., & García, M. (2007). *Aprendizaje y desigualdad social en México*. México D. F.: INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN.
- Backhoff Escudero, E., Bouzas Riaño, A., Hernández, E., & García, M. (2007). *Factores escolares y aprendizaje en México (El caso de la educación básica)*. México: INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN.
- Barro, R. J. (2013). Education and Economic Growth. En *ANNALS OF ECONOMICS AND FINANCE* (págs. 280 -290). Harvard University.
- Bowles, S., & Gintis, H. (1985). *La instrucción escolar en la América capitalista*. México D. F.
- Braslavsky, C., & Cosse, G. (2003). *PANORAMA INTERNACIONAL SOBRE CALIDAD Y EQUIDAD EN LA EDUCACION*. Congreso "Calidad, Equidad y Educación".
- Castañeda, D. (16 Mayo 2013). Educación y calidad de vida, tarea pendiente en México. *Forbes Magazine online*.
- COPREDEH. (2011). Artículos 13 Y 14. En COPREDEH, *Pacto Internacional de los Derechos Económicos, Sociales y Culturales*. Guatemala.
- Hanushek, E. A., & Wößmann, L. (Febrero 2017). *The Role of Education Quality in Economic Growth*. World Bank Policy Research Working Paper 4122.
- Hopkins, J., & Coleman, J. (1966). *Equality of educational opportunity*. Washington, D.C.: Government Printing.

- INEE. (2005). *Excale, Exámenes de la calidad y el logro Educativo, proceso de construcción y características básicas. Los temas de la evaluación, colección de folletos*. México: INEE.
- INEE. (2012). *La Educación en México: Estado actual y consideraciones sobre su evaluación*. Ciudad de México: Presentación del INEE ante la Comisión de Educación de la LXII Legislatura de la Cámara de Senadores.
- INEE. (2012). *Panorama Educativo de México*. México.
- Lagunes, R. V., & Maldonado Márquez, J. R. (2016). *LA EVOLUCIÓN DE LA CALIDAD EDUCATIVA*. MÉXICO, D.F.: UNIVERSIDAD PANAMERICANA.
- Martínez Trinidad, J. F., & Carrasco Ochoa, J. (2015). Reconocimiento de Patrones. *Komputer Sapiens, Revista de Divulgación de la Sociedad Mexicana de Inteligencia*, 3, 5-10.
- Muñoz Izquierdo, C., & Guzmán, J. T. (1971). Una exploración de los factores determinantes del rendimiento escolar en la educación primaria. *Revista Latinoamericana de Estudios Educativos*, 167-191.
- Muñoz Izquierdo, C., Márquez, A., Sandoval, A., & Sánchez, H. (2004). *Factores Externos e Internos a las Escuelas que Influyen en el Logro Académico de los Estudiantes de Nivel Primaria en México, 1998–2002. Análisis Comparativo Entre Entidades con Diferente Nivel de Desarrollo*. México: Instituto de Investigaciones para el Desarrollo de la Educación Universidad Iberoamericana, Ciudad de México.
- Ochoa Somuano, J. (2005). *Técnicas de Selección de Atributos para la Categorización Automática de Escenas Visuales*. Cuernavaca, Morelos: Maestría, Centro Nacional de Investigación y Desarrollo Tecnológico.
- Robles Vásquez, H. V., & Pérez Miranda, M. G. (2018). *PANORAMA EDUCATIVO DE MÉXICO (Indicadores del Sistema Educativo Nacional 2017)*. Ciudad de México.: Blanca Gayosso Sánchez.
- Ruíz Shulcloper, J., Alba Cabrera, E., & Lazo Cortés, M. (1995). *Introducción a la teoría de Testores*. México, DF: Departamento de Ingeniería Electrica, CINVESTAV-IPN.
- Ruíz Shulcloper, J. (2013). Acerca del surgimiento del Reconocimiento de Patrones en Cuba. *Revista Cubana de Ciencias Informáticas*, 7.
- Sánchez Garfias, F., Díaz de León Santiago, J., & Yáñez Márquez, C. (2003). Reconocimiento automático de patrones: conceptos básicos. *Research in Computer Science*, 10, 91-102.
- Santiesteban Alganza, Y., & Pons Porrata, A. (2003). LEX: Un nuevo algoritmo para el cálculo de los testores típicos. *Revista Ciencias Matemáticas*, 21, 85-95.

Simon, S. (1995). Acceso y retrasos en la educación en América Latina. *publicaciones SITEAL*, 2.

UNESCO. (2005). *Informe de seguimiento del programa educación para todos 2005*. París, Francia: UNESO Publications.

Yáñez, C. (2002). Memorias Asociativas Basadas en Relaciones de Orden y Operaciones. *Computación y Sistemas*, 6, 300-311.

Zamudio Mesa, C., Díaz Argüero, C., & Lepe García, E. (2012). *El aprendizaje de los contenidos curriculares de Español Un análisis de los resultados de los Exámenes de la Calidad y el Logro Educativos (Excale 03, 06 y 09)*. México: INEE.

