



**UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES**

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

CENTRO DE CIENCIAS BÁSICAS

**Creación y Puesta en Marcha de Algoritmo Metaheurístico Híbrido
de Aprendizaje Automático para el Reconocimiento de Factores en
Diabetes Mellitus.**

TESIS

QUE PARA OBTENER EL GRADO DE
DOCTOR EN CIENCIAS APLICADAS Y TECNOLOGÍA

Presenta:

MITC. PABLO RODRIGUEZ DE LEON

Tutora:

DRA. MARÍA DOLORES TORRES SOTO

Co-tutora:

DRA. AURORA TORRES SOTO

Comité tutorial:

DRA. EUNICE ESTHER PONCE DE LEÓN SENTI

Aguascalientes, Ags. 23 de Noviembre de 2022.

CARTA DE VOTO APROBATORIO
INDIVIDUAL

MTRO. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTORA* designada del estudiante *PABLO RODRIGUEZ DE LEÓN* con ID *173310* quien realizó la tesis titulada: *CREACIÓN Y PUESTA EN MARCHA DE ALGORITMO METAHEURÍSTICO HÍBRIDO DE APRENDIZAJE AUTOMÁTICO PARA EL RECONOCIMIENTO DE FACTORES EN DIABETES MELLITUS*, 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 imprimirla 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 18 de Noviembre de 2022.



DRA. MARÍA DOLORES TORRES SOTO
Tutora de tesis

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

Elaborado por: Depto. Apoyo al Posgrado.
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-07
Actualización: 01
Emisión: 17/05/19

CARTA DE VOTO APROBATORIO
INDIVIDUAL

MTRO. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTORA* designada del estudiante **PABLO RODRIGUEZ DE LEÓN** con ID **173310** quien realizó la tesis titulada: **CREACIÓN Y PUESTA EN MARCHA DE ALGORITMO METAHEURÍSTICO HÍBRIDO DE APRENDIZAJE AUTOMÁTICO PARA EL RECONOCIMIENTO DE FACTORES EN DIABETES MELLITUS**, 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 imprimirla 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 18 de Noviembre de 2022.


DRA. AURÓRA TORRES SOTO
Co-Tutora de tesis

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

Elaborado por: Depto. Apoyo al Posgrado.
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-07
Actualización: 01
Emisión: 17/05/19



CARTA DE VOTO APROBATORIO
INDIVIDUAL

MTRO. EN C. JORGE MARTÍN ALFÉREZ CHÁVEZ
DECANO (A) DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTOR* designado del estudiante **PABLO RODRIGUEZ DE LEÓN** con ID **173310** quien realizó la tesis titulada: **CREACIÓN Y PUESTA EN MARCHA DE ALGORITMO METAHEURÍSTICO HÍBRIDO DE APRENDIZAJE AUTOMÁTICO PARA EL RECONOCIMIENTO DE FACTORES EN DIABETES MELLITUS**, 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 imprimirla 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 18 de Noviembre de 2022.

DRA. EUNICE ESTHER PONCE DE LEÓN SENTI
Asesora de tesis

El nombre completo que aparece en el Voto Aprobatorio debe coincidir con el que aparece en el documento empastado. No se puede abreviar, ni omitir nombres

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

Elaborado por: Depto. Apoyo al Posgrado.
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-07
Actualización: 01
Emisión: 17/05/19



DICTAMEN DE LIBERACIÓN ACADÉMICA PARA INICIAR LOS TRÁMITES DEL EXAMEN DE GRADO



Fecha de dictaminación dd/mm/aaaa: 24/11/2022

NOMBRE: Pablo Rodríguez de León **ID** 173310

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

TIPO DE TRABAJO: () Tesis () Trabajo Práctico
TÍTULO: Creación y Puesta en Marcha de Algoritmo Metaheurístico Híbrido de Aprendizaje Automático para el Reconocimiento de Factores en Diabetes Mellitus

IMPACTO SOCIAL (señalar el impacto logrado): Mediante el desarrollo e implementación de un algoritmo de testeros típicos se realizó la propuesta de reagrupación de unidades de salud respecto a sus jurisdicciones de responsabilidad y se encontraron las zonas geográficas con mayor incidencia de diabetes mellitus, traducándose esto en una detección temprana de la Diabetes Mellitus.

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

INDICAR	SI	NO	N.A. (NO APLICA)	SEGÚN CORRESPONDA:
Elementos para la revisión académica del trabajo de tesis o trabajo práctico:				
SI				El trabajo es congruente con las LGAC del programa de posgrado
SI				La problemática fue abordada desde un enfoque multidisciplinario
SI				Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI				Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI				Los resultados presentados en el trabajo son de gran relevancia científica, 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
NO				Generó transferencia del conocimiento o tecnológica
SI				Cumple con la ética para la investigación (reporte de la herramienta antiplagio)
El egresado cumple con lo siguiente:				
SI				Cumple con lo señalado por el Reglamento General de Docencia
SI				Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc)
SI				Cuenta con los votos aprobatorios del comité 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
SI				Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)
En caso de Tesis por artículos científicos publicados				
N.A.				Aceptación o Publicación de los 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: SI No

Elaboró: **FIRMAS**

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

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

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

Revisó:

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

Autorizó:

NOMBRE Y FIRMA DEL DECANO: Mtro. Jorge Martín Alférez Chávez

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

RESEARCH IN COMPUTING SCIENCE

ISSN 1870-4069

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

Mexico City, July 1st, 2021

To whom it may concern:

Hereby I confirm that the paper

“Propuesta de Distribución y Agrupamiento de Centros de Salud en el Estado de Aguascalientes Basado en Algoritmo K-Means”

by Pablo Rodríguez de León, María Dolores Torres Soto and Aurora Torres Soto

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

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

With best regards,



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

RESEARCH IN COMPUTING SCIENCE

ISSN 1870-4069

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

Mexico City, September 1st, 2021

To whom it may concern:

Hereby I confirm that the paper

“Proposals for Grouping Health Centers and Health Jurisdictions in the State of Aguascalientes Using Haversine Distance and K-Means Algorithm”

by Pablo Rodríguez de León, María Dolores Torres Soto, Aurora Torres Soto and Eunice Esther Ponce de León Sentí.

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

It is scheduled for the volume 150(9), 2021, which is now in the process of technical production.

With best regards,



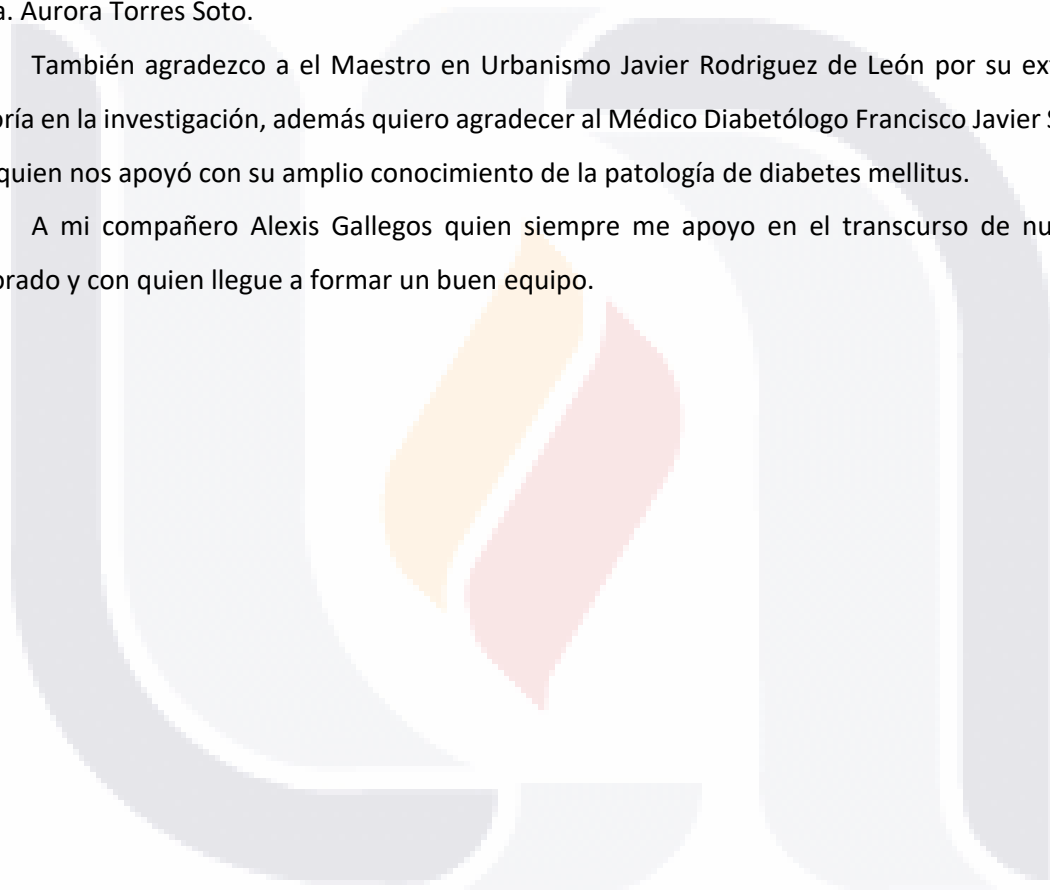
.....
Dr. Grigori Sidorov
Editor-in-Chief

AGRADECIMIENTOS

Aprovecho este espacio para agradecer el apoyo de todas las personas que hicieron fuera posible culminar esta etapa de mi vida académica en ayudarme y darme ánimos en momentos en que creía que este trabajo no acabaría en especial a mis tutoras la Dra. María Dolores Torres Soto y la Dra. Aurora Torres Soto.

También agradezco a el Maestro en Urbanismo Javier Rodriguez de León por su extensa asesoría en la investigación, además quiero agradecer al Médico Diabetólogo Francisco Javier Serna Vela quien nos apoyó con su amplio conocimiento de la patología de diabetes mellitus.

A mi compañero Alexis Gallegos quien siempre me apoyo en el transcurso de nuestro doctorado y con quien llegue a formar un buen equipo.



DEDICATORIA

Este trabajo está dedicado a la memoria de mis padres, Silvia Isabel María de Lourdes de León García y Pablo Rodríguez Portales, ya que ellos me apoyaron toda su vida e impulsaron y motivaron el esfuerzo y dedicación para que yo realizara esta meta y sueño.



ÍNDICE DE CONTENIDO

Resumen.....	10
Abstract	11
Capítulo I. Introducción.....	12
1.1. Introducción.	12
1.2. Problema de Investigación.	12
1.3. Hipótesis.....	13
1.4. Objetivos.	13
1.5. Antecedentes.	14
1.6. Estructura de la tesis.	15
Capítulo II. Marco Teórico.....	17
2.1. Aprendizaje Automático.....	17
2.2. Heurísticas.....	18
2.3. Metaheurísticas.....	20
2.4. Testores.....	22
2.4.1 Método de búsqueda.....	23
2.4.2 Reglas de comparación.	24
2.2.3 Ejemplo.....	27
2.5. Redes Neuronales Artificiales.....	32
2.6 Deep Learning	35
2.6.1 Arquitectura y Características	36
2.6.2 Entrenamiento en Redes Neuronales	38
2.6.3 Función de Activación.	39
2.6.4 Tipos de Redes Neuronales más Comunes.	42
2.7. Clasificadores Bayesianos.....	43

2.7.1 Naïve Bayes.	46
2.8. Diabetes Mellitus.	46
2.9. K-Means.....	52
2.10. CIE 10.....	53
2.11. Métodos de decisión multicriterio.....	59
2.12. Técnica de Preferencia de Orden por Similitud a la Solución Ideal.	59
Capítulo III. Metodología de la investigación	62
3.1. Introducción.	62
3.2. Objeto de Estudio.....	62
3.3. Criterios de Selección	63
3.3.1. Criterios de inclusión.....	63
3.3.2. Criterios de exclusión.	63
3.4. Definición de Variables.....	64
3.4. 1 Definición Constructiva.	65
3.5 Cálculo del Tamaño de la Muestra.....	67
Capítulo IV. Marco de trabajo para creación y puesta en marcha de sistema metaheurístico de aprendizaje automático para factores de diabetes mellitus	68
4.1. Marco de trabajo de algoritmo metaheurístico de aprendizaje automático.	71
4.2. Preparación de la información.	72
Capítulo V. Implementación del Sistema SMAADM.....	79
Capítulo VI. Propuesta de reubicación de jurisdicciones	84
6.1..... Diseño e implementación de algoritmo K-Means modificado	84
6.2 Diseño e implementación de TOPSIS	91
Capítulo VII. Productos y Actividades derivados de la investigación	96
7.1. Convenio de colaboración UAA-ISSEA.....	96

7.2. Producto de selección de subconjuntos de características de testores típicos..... 96

7.3. Establecimiento de parámetros para redes neuronales 98

7.4. Producto clasificador Ingenuo de Bayes 100

Capítulo VIII. Discusión y conclusiones 102

Capítulo IX Referencias. 106

Capítulo X Anexo. 112

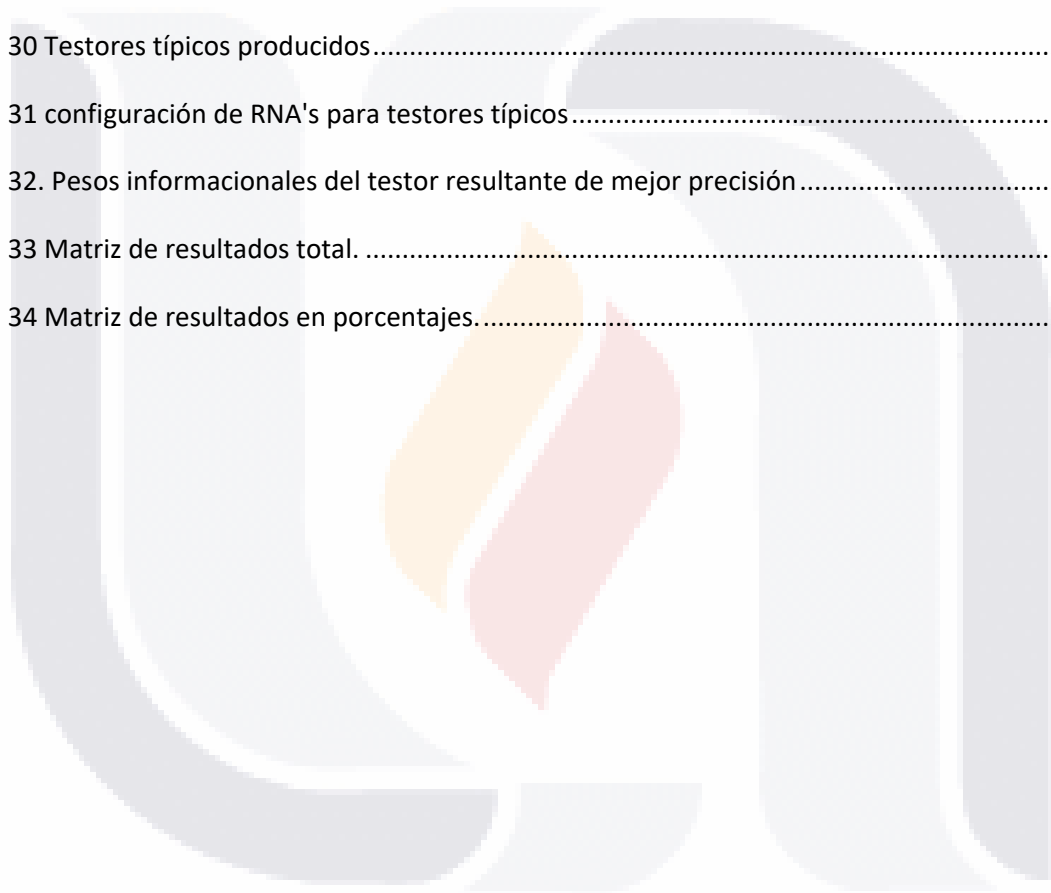
10.1 Convenio de colaboración..... 112



ÍNDICE DE TABLAS

Tabla 1 Ejemplo de Variables y sus Componentes de Análisis.....	27
Tabla 2 Ejemplo de matriz de Aprendizaje.....	28
Tabla 3 Ejemplo de Matriz de Diferencias y Matriz Básica	29
Tabla 4 Ejemplo de Exploración y Conclusión de Candidatos	30
Tabla 5 Ejemplo de Testores Encontrados	32
Tabla 6 Ejemplo de Peso Informacional	32
Tabla 7 Incidencia en México de Diabetes Mellitus Grupo de Edad y Sexo.....	50
Tabla 8 Variables Iniciales de la Base de Datos	64
Tabla 9 Clasificación de Etapas de la Vida.....	72
Tabla 10 Discretización de Edad.....	73
Tabla 11 Discretización de Glucemia	73
Tabla 12 Clasificación de Hipertensión Arterial	74
Tabla 13 Discretización de Tensión Arterial.....	74
Tabla 14 Frecuencia Cardíaca.....	75
Tabla 15 Discretización de Frecuencia	75
Tabla 16 : Escala de Temperatura Corporal	76
Tabla 17 Discretización de Temperatura Corporal.....	77
Tabla 18 Dominio y Criterio de Comparación de Variables	78
Tabla 19 Distribución Actual de Población.....	88
Tabla 20 Distribución de Población Propuesta	89
Tabla 21 Posibles localidades	91
Tabla 22 Matriz de criterios de decisión 1	92
Tabla 23 Matriz de criterios de decisión 2	92

Tabla 24 Matriz normalizada 1.....	93
Tabla 25 Matriz normalizada 2.....	93
Tabla 26 Matriz de solución anti ideal positiva 1.....	94
Tabla 27 Matriz de solución anti ideal positiva 2.....	94
Tabla 28 Matriz de proximidad relativa a la solución absoluta	94
Tabla 29 Lista de clasificación de soluciones	95
Tabla 30 Testores típicos producidos.....	97
Tabla 31 configuración de RNA's para testores típicos.....	99
Tabla 32. Pesos informacionales del testor resultante de mejor precisión.....	99
Tabla 33 Matriz de resultados total.	101
Tabla 34 Matriz de resultados en porcentajes.....	101



ÍNDICE DE FIGURAS

Figura 1 Elementos de RNA Cita.....	34
Figura 2 Elementos de Neurona Artificial	35
Figura 3 de Inteligencia Artificial a Deep Learning.....	36
Figura 4 Función lineal	40
Figura 5 Función Sigmoide	41
Figura 6 Función Tangente Hiperbólica.....	41
Figura 7 DNN	42
Figura 8 CNN.....	43
Figura 9 RNN.....	43
Figura 10 Incidencia en México de Diabetes Mellitus Grupo de Edad y Sexo	49
Figura 11 Casos de Hospitalización por Diabetes Mellitus por Grupo de Edad	51
Figura 12 Marco de Trabajo de la investigación	70
Figura 13 Marco de Trabajo de algoritmo metaheurístico de aprendizaje automático.....	71
Figura 14 implementación del sistema SMAADM.....	80
Figura 15 Alerta de considerar paciente diabético	81
Figura 16 Alerta o sugerencia de considerar paciente prediabético	82
Figura 17 Alerta o sugerencia de considerar paciente no diabético.....	83
Figura 18 Procedimiento de reubicación de jurisdicciones y reagrupación de unidades de salud ..	84
Figura 19 Localización de posible ubicación PDA 1.....	86
Figura 20 Localización de posible ubicación PDA 2.....	87
Figura 21 Localización de posible ubicación PDA 3.....	87
Figura 22 Localización de posible ubicación PDA 4.....	88
Figura 23 Distribución y ubicación actuales de centros de salud ISSEA.....	89

Figura 24 Nueva agrupación propuesta de centros de salud ISSEA..... 90



ÍNDICE DE ECUACIONES

Ecuación 1	26
Ecuación 2	28
Ecuación 3	28
Ecuación 4	28
Ecuación 5	28
Ecuación 6	40
Ecuación 7	40
Ecuación 8	41
Ecuación 9	44
Ecuación 10	44
Ecuación 11	45
Ecuación 12	45
Ecuación 13	45
Ecuación 14	46
Ecuación 15	60
Ecuación 16	60
Ecuación 17	61
Ecuación 18	61
Ecuación 19	61
Ecuación 20	61
Ecuación 21	61
Ecuación 22	67
Ecuación 23	67

Ecuación 24 98



Resumen.

Cada año se pierden muchas vidas en el mundo y nuestro país debido a la diabetes mellitus, la cual es una enfermedad crónica que es considerada junto con la hipertensión y la obesidad se considera como una de las más impactantes en nuestro país, de ahí la importancia de investigar los factores que determinan cómo se desarrolla esta patología. Mediante esta investigación, se diseñó, desarrolló y creó un algoritmo metaheurístico híbrido de aprendizaje automático para identificar temprana y oportuna de pacientes predispuestos a padecer esta enfermedad en la población de Aguascalientes. En esta tesis se desarrolló un marco de trabajo el cual se describe a continuación:

Este marco de trabajo se realizó con información real de 91 bases de datos, 86 pertenecientes a centros de salud y 5 pertenecientes a hospitales generales del *Instituto de Servicios de Salud del Estado de Aguascalientes*, dichas bases de datos fueron procesadas filtradas y discretizadas para su procesamiento en los siguientes estudios.

Mediante el desarrollo e implementación de un algoritmo de testores típicos se trabajó una línea de investigación con la cual se obtuvieron cinco testores típicos, con las variables de los testores típicos resultantes se diseñaron y desarrollaron 5 redes neuronales y paralelamente se desarrolló e implemento un clasificador ingenuo de bayes, con el propósito de comparar y obtener la mejor precisión de los clasificadores de redes neuronales desarrollados y posteriormente comparar esa precisión con la precisión del ingenuo de bayes y finalizar seleccionando el mejor clasificador.

Además se realizó la propuesta de reagrupación de unidades de salud respecto a sus jurisdicciones de responsabilidad y se encontraron las zonas geográficas con mayor incidencia de diabetes mellitus, mediante el desarrollo y aplicación de un algoritmo K-means modificado se propone una nueva jurisdicción sanitaria con la finalidad de mejorar la distribución y reducción de distancia de los centros de salud hacia las jurisdicciones de responsabilidad, traduciéndose esto en una mejor atención, mejor distribución de los insumos para los centros de salud e incluso una detección temprana de patologías tan importantes como la Diabetes Mellitus.

Cabe resaltar que los resultados de esta investigación además de ser novedosos proporcionan información relevante los cuales dan la pauta para investigaciones futuras de diabetes mellitus y otras patologías relevantes en la población de Aguascalientes.

Abstract.

Every year many lives are lost worldwide and in our country because of diabetes mellitus, which is a chronic disease that together with hypertension and obesity is considered one of the most impacting diseases in our country, hence the importance of investigating the factors that determine how this pathology develops. Through this research, a hybrid metaheuristic machine learning algorithm was designed, developed and created to identify early and opportune patients predisposed to develop this disease in the population of Aguascalientes. In this thesis a framework was developed which is described below:

This framework was performed with real information from 91 databases, 86 belonging to health centers and 5 belonging to general hospitals of the Instituto de Servicios de Salud del Estado de Aguascalientes, these databases were processed filtered and discretized for processing in the following studies.

Through the development and implementation of an algorithm of typical testers, a line of research was worked with which five typical testers were obtained, with the variables of the resulting typical testers 5 neural networks were designed and developed and in parallel a naive Bayes classifier was developed and implemented, with the purpose of comparing and obtaining the best accuracy of the developed neural network classifiers and then comparing that accuracy with the accuracy of the naive Bayes classifier and finally selecting the best classifier.

In addition, a proposal was made to cluster health units with respect to their jurisdictions of responsibility and the geographic areas with the highest incidence of diabetes mellitus were found. Through the development and application of a modified K-means algorithm, a new health jurisdiction was proposed with the purpose of improving the distribution and reducing the distance from the health centers to the jurisdictions of responsibility, resulting in better care, better distribution of supplies for the health centers and even early detection of pathologies as important as Diabetes Mellitus.

It should be highlighted that the results of this research, in addition to being novel, provide relevant information that will set the guidelines for future research on diabetes mellitus and other relevant pathologies in the population of Aguascalientes.

CAPÍTULO I INTRODUCCIÓN

Capítulo I. Introducción

1.1. Introducción.

Día a día se pierden muchas vidas debido a la diabetes mellitus, la cual es una enfermedad crónica vinculada a otras enfermedades como hipertensión, dislipidemia y síndrome metabólico, por esta razón, emerge la inquietud de identificar anticipadamente las variables características relacionadas con el desarrollo de esta enfermedad.

En esta investigación, se crea un marco de trabajo que contiene el desarrollo de un algoritmo metaheurístico híbrido de aprendizaje automático para la obtención de datos relevantes que serán extraídos de bases de datos de pacientes con la patología de Diabetes Mellitus del Instituto de Servicios de Salud del Estado de Aguascalientes (ISSEA).

Se hará uso de la clasificación establecida por el CIE10 (Panamericana, 2003), además se contó con la asesoría del médico diabetólogo Francisco Javier Serna Vela.

Los datos corresponden a la colección de información de 89 centros de salud y 5 hospitales, los cuales se obtuvieron por medio de un convenio de cooperación de información establecido con el Instituto de Salud del Estado de Aguascalientes (ISSEA).

1.2. Problema de Investigación.

El problema, desde el punto de vista computacional, es el reconocimiento de patrones aplicado a la identificación de factores de riesgo de diabetes mellitus.

La Diabetes Mellitus es una de las principales causas de muerte (FID, 2015) al grado que se ha llegado a considerar una epidemia.

El diagnóstico médico es complicado y subjetivo. Su resultado no depende de una solución sistematizada sino de la interpretación de los síntomas del paciente.

CAPÍTULO I INTRODUCCIÓN

Existen razones ideológicas y culturales por las cuales se cree que una computadora no es capaz de diagnosticar como un médico humano, además existe resistencia de algunos profesionales de la salud a sentirse cuestionados, observados o aconsejados por una máquina.

1.3.Hipótesis.

1. Los factores de pacientes con la patología de diabetes mellitus serán reconocidos por el procesamiento de información por medio del algoritmo metaheurístico híbrido de aprendizaje automático.
2. Existen focos rojos geográficos de concentración de pacientes con la patología de diabetes mellitus en el estado de Aguascalientes.
3. Mediante un algoritmo metaheurístico híbrido se identificarán los factores en los diversos grupos de Diabetes Mellitus clasificados por clave CIE10.
4. La hipertensión en conjunto con el incremento en glucemia, los cambios en el índice de masa corporal y el género no son los únicos factores importantes en la determinación de la patología de diabetes mellitus.

1.4. Objetivos.

Objetivo General

- Diseño e implementación de un algoritmo metaheurístico híbrido de aprendizaje automático para identificar los factores de riesgo asociados con la patología de Diabetes Mellitus, así como identificar las zonas geográficas de mayor incidencia de dicha patología en el estado de Aguascalientes

Objetivos Específicos

- Diseñar e implementar un marco de trabajo para el diseño, implementación y puesta a punto de un algoritmo metaheurístico híbrido, así como su validación para Identificar los factores críticos en los pacientes de la población de Aguascalientes diagnosticados con Diabetes Mellitus.
- Identificar los centros geográficos de concentración de pacientes con la patología de diabetes mellitus para la mejor aplicación de políticas públicas.
- Proponer una distribución de centros de salud para mejorar los tiempos de traslado de personal y medicamentos, detección de atención temprana de diabetes mellitus, reducir

CAPÍTULO I INTRODUCCIÓN

costos de traslados y tener una mejor distribución por jurisdicción de pacientes totales con diabetes.

1.5. Antecedentes.

En el campo de la medicina, es delicado realizar el diagnóstico de una enfermedad, uno de los inconvenientes más importantes de este proceso, es la subjetividad del especialista que lo efectúa. Este hecho se hace notar en particular en actividades de reconocimiento de patrones y características, donde la experiencia del profesional está directamente relacionada con el diagnóstico final. Lo anterior, es debido al hecho de que el resultado no depende de una solución sistematizada sino de la interpretación de los síntomas del paciente

La diabetes Mellitus, es la octava causa de muerte en el mundo del año 2000 al 2012; así mismo, el INEGI sitúa a la Diabetes Mellitus como una de las 3 primeras causas de muerte desde 2006 a nivel nacional y como segunda causa de muerte a nivel nacional en 2018 (INEGI, 2019). El Gobierno Estatal de Aguascalientes la posiciona como la segunda causa en la mortalidad general con un 14.67% del total de las defunciones en 2018 (Gobierno del Estado de Aguascalientes, 2019).

Como se conoce, esta patología no discrimina condición social, edad o género, de ahí la relevancia de identificar las características importantes que describen cómo se desarrolla esta enfermedad, que junto con la hipertensión es considerada como una de las más impactantes en nuestro país. Mediante esta investigación, se busca encontrar los factores de riesgo en la población de Aguascalientes.

Como mencionan (Alvarado-guevara & Flores-sandí, 2009), El diagnóstico de patologías forma parte de las cinco áreas de interés del cuidado médico del paciente (tratamiento médico y quirúrgico, diagnóstico, comunicación de información/documentación, comunicación interpersonal y otros de tipo administrativo) pueden listarse con número. El diagnóstico de patologías es uno de los momentos más importantes de la actividad médica; constituye el acto inicial de la relación del profesional con el paciente, su finalidad es identificar y determinar la naturaleza y las causas de la patología. De él depende la elección del tratamiento adecuado y el momento cuando se indica al

CAPÍTULO I INTRODUCCIÓN

paciente cuál es la enfermedad o mal que lo afecta, para después brindar la información necesaria. (Lugo-Reyes¹, Maldonado-Colín², & Murata³, 2014)

El origen de los errores médicos puede vincularse con la ejecución y disposición del servicio médico otorgado ya sea por:

- a) Exceso de servicio, sobre uso de medicamento o tratamiento, el exceso de servicio provoca que el médico no esté en óptimas condiciones debido al exceso de horas laborales y también se refiere al uso continuo de un tratamiento específico en un grupo determinado de pacientes sin previos análisis.
- b) Insuficiencia de servicio, ausencia o reducido uso de un medicamento o tratamiento, la ausencia de servicio es cuando el médico no cuenta con la suficiente experiencia en consulta y esto provoca caer en diagnósticos médicos erróneos, la ausencia o reducido uso de medicamentos o tratamientos se refiere al poco uso de medicamentos de una patología según su criterio.
- c) Diagnóstico inapropiado, es el uso de un diagnóstico que no corresponde a la patología que presenta el paciente.

El diagnóstico inapropiado es un problema progresivo en los sistemas de salud, dado que se registra un incremento en las demandas por parte de los pacientes, lo cual se adjudica a esta clase de incidentes. Según estudios realizados en la Unión Europea y Estados Unidos aproximadamente un 10% de los ingresos hospitalarios implican algún tipo de daño causado al paciente por un diagnóstico inapropiado. (Woolf, Kuzel, Dovey, & Phillips, 2004)

Un dato interesante en el diagnóstico médico es la tasa de diagnósticos incorrectos en la práctica clínica como lo mencionan (Lugo-Reyes¹ et al., 2014) la cual depende de la fuente y el escenario; "así, se ha estimado en 150 de 1,000 pacientes, lo cual es del 10 al 20% (en el servicio de urgencias y en estudios de discrepancias con autopsias). "Errar es humano" y en los hospitales causa malestar, daños, gastos y muertes, especialmente en diagnósticos serios como melanomas, hipertensión o diabetes". (Reed, May, Contributors, & Nicholas, 2011)

1.6. Estructura de la tesis.

Esta tesis se compone de diez capítulos y un anexo descritos a continuación:

CAPÍTULO I INTRODUCCIÓN

1. Introducción, en este capítulo se abordan el problema de investigación, una breve reseña de la necesidad e importancia de la investigación, el objetivo general, los objetivos específicos y las hipótesis de la investigación.
2. Marco Teórico, en este se presenta la información y conceptos referentes a los fundamentos necesarios para la comprensión del lector, abordando temas como diabetes mellitus, heurísticas, testores típicos, redes neuronales, K-means, distancia Haversine, etc.
3. Metodología de la Investigación., aquí se presenta el objeto de estudio, así como los criterios de selección y definición de variables utilizadas en la investigación, se presenta el marco de trabajo que guía la presente investigación.
4. Marco de trabajo para la creación y puesta en marcha del sistema metaheurístico de aprendizaje automático para factores de diabetes mellitus (SMAADM), aquí se prepara la información.
5. Implementación del SMAADM, se presenta los pasos de la implementación, así como pantallas del sistema en funcionamiento.
6. Propuesta de reubicación de jurisdicciones, en este capítulo se presentan los pasos de la reagrupación de centros de salud con respecto a sus jurisdicciones de responsabilidad y la nueva ubicación de una propuesta de una nueva jurisdicción.
7. Productos y actividades derivados de la investigación, aquí se presentan productos derivados de importancia en la investigación.
8. Conclusiones y Discusión, en este capítulo, se presentan las conclusiones obtenidas en el punto de finalización de este trabajo, conclusiones a los objetivos de investigación generales y específicos de la investigación.
9. Bibliografía.
10. Anexo.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Capítulo II. Marco Teórico.

En esta sección, se asentarán los fundamentos que deben ser comprendidos por el lector para identificar el ámbito de las aportaciones de esta tesis, se hablará acerca de heurísticas, como algunos de problemas de optimización que enfrenta el área de la inteligencia artificial y cómo mediante el uso de metaheurísticas se puede atender el problema indicado. Se presentan también mecanismos de selección de características, se hace una breve introducción al concepto de testores, redes neuronales, ingenuo de Bayes, K-means, distancia Haversine, TOPSIS e igualmente se expone una breve introducción al diagnóstico de diabetes Mellitus, sus características y factores de riesgo para una mejor comprensión del tema.

2.1. Aprendizaje Automático.

El aprendizaje automático, (Machine Learning o también llamado aprendizaje artificial) (Plaza, 1992) es la parte de la Inteligencia Artificial que tiene como finalidad producir técnicas que permitan a los sistemas aprender. Es el diseño, creación y puesta en marcha de algoritmos capaces de reconocer patrones y concretar conductas fundamentadas en información suministrada en forma de ejemplos o modelos.

Se puede decir que el aprendizaje automático es un tipo de aprendizaje capaz de reconocer la situación y responder con la habilidad aprendida; (como el ser humano), mediante mecanismos de resolución con los cuales es capaz de mejorar su acción en diversas tareas, como resultado de haber solucionado un problema específico.

Tom Mitchell escribió en términos operativos: *“Se dice que una computadora aprende de la experiencia (E) con respecto a alguna clase de tareas (T) y medida de desempeño (P), si su desempeño en las tareas T, medida mediante P, mejora con la experiencia E”*. (Mitchell, 1983)

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

La importancia del desarrollo de sistemas de aprendizaje automático se basa en el hecho de que en muchos entornos se carece de la experiencia, comúnmente la representación del conocimiento es limitada, incompleta o nula. Por otro lado, el generar elementos o agentes dotados de todo el conocimiento requerido con un mínimo de errores es tardado, complejo y costoso.

Por mencionar algunas investigaciones de aprendizaje automático en medicina a través de los años se encuentra el Homer Warner en 1961 el cual desarrolló un sistema de diagnóstico automatizado para diagnosticar específicamente problemas cardiológicos aplicado a 35 pacientes con cardiopatías congénitas a partir de 50 variables clínicas. Homer Warner obtuvo un resultado superior al de los cardiólogos referentes. (Homer R. Warner, M.D., Ph.D.; Alan F. Toronto, M.D.; L. George Veasey, 1961)

En la década de los 70's, la universidad de Stamford desarrolló un sistema experto llamado MYCIN que no se puso en práctica por cuestiones legales, pero que, sin embargo, ayudaba al diagnóstico clínico de infecciones y trastornos de coagulación en la sangre. Este sistema obtuvo un 65% de efectividad en tratamientos diagnosticados correctamente, lo cual demostró ser superior que 5 profesionales humanos expertos en infectología. (Cleophas, Ton J., Zwinderman, 2014)

En los 80's se desarrolló una herramienta electrónica educativa para asistencia en la toma de decisiones en medicina interna llamada CADUCEUS/INTERNIST-1/QUICK MEDICAL REFERENCE, la cual fue diseñada para diagnosticar 1000 enfermedades usando un motor de inferencia parecido al de MYCIN, utilizando un grafo acíclico direccionado que establecía relaciones de causalidad entre las variables. Esta herramienta no obtuvo éxito debido al excesivo tiempo que tomaba la consulta y la dificultad para utilizar la interfaz de usuario. (Siegel & Parrino, 1988).

2.2. Heurísticas.

Una definición establece que: las heurísticas son técnicas que buscan soluciones de buena calidad (de valor cercano al óptimo) a un costo computacional justo, aunque sin garantizar la optimalidad de esta. En general, no se reconoce el grado de error. Pueden ser agrupadas de manera muy general en la siguiente clasificación. (Rodríguez Bocca, 2008)

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Métodos constructivos: los cuales generan una solución y van agregando elementos de esta hasta completarla.

Métodos de búsqueda local: Los cuales comienzan desde una solución inicial y repetidamente van sustituyendo la solución más cercana y de mejor calidad.

El método heurístico por excelencia se refiere a un mecanismo voraz o greedy el cual construye una solución eligiendo de manera iterativa los elementos o piezas de menor costo o dirigiéndose al vecino mejor calificado. Para algunos problemas con estructura particular, la solución encontrada es la solución óptima, pero en general, no es el caso.

Actualmente se recomienda la hibridación de mecanismos heurísticos, ya que estos producen mejores soluciones que los mecanismos puristas (Torres Soto, 2010) y (Raidl, 2006).

En años recientes, se han presentado un gran número de algoritmos que no siguen los conceptos puristas de una metaheurística tradicional simple, sino que éstos son la combinación de varios algoritmos, los cuales fuera del campo tradicional de las metaheurísticas, son comúnmente llamados metaheurísticas híbridas (Raidl, 2006).

Para comprender un poco más la interpretación de metaheurística híbrida, podemos mencionar el significado de híbrido:

- Que proviene de 2 o más componentes de naturaleza distinta.
- Que tiene progenitores de diferente especie.
- Obtenido del cruce de dos individuos de distinta especie
- En general, que está formado por elementos de distinta naturaleza

Debido al auge o creciente desarrollo de aplicaciones de metaheurísticas híbridas así como los workshops de metaheurísticas híbridas (Blum & Roli, 2004), (Blesa, Blum, Roli, & Sampels, 2005) documentan la popularidad, éxito e importancia de esta línea de investigación. De hecho, hoy parece que escoger la adecuada metodología híbrida es decisivo para lograr el mejor desempeño en la solución de problemas difíciles.

Al principio, la idea de hibridar no fue muy popular, ya que los grupos de investigación estaban alejados, y cada uno consideró su propio desarrollo como el mejor siguiendo sus filosofías.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

A continuación, una representación de la clasificación resumida de metaheurísticas híbridas, donde se establecen las condiciones de la hibridación que se propone:

- a) se puede hibridar una metaheurística con otra.
- b) Una metaheurística con algoritmo específico para el problema que se está considerando, como simulaciones especiales.
- c) Metaheurísticas con otras técnicas generales que vienen de campos tales como investigación de operaciones o inteligencia artificial, son ejemplos exactos como los de ramificación y poda (Narendra & Fukunaga, 1977) (Raidl, 2006).

Otra propiedad para distinguir los sistemas híbridos es el orden de la ejecución.

- 1) El modelo secuencial o de lotes, donde un algoritmo se realiza estrictamente después de otro y solo pasa información en una dirección.
- 2) En el modelo intercalado, en el cual no se respeta un orden secuencial entre los algoritmos involucrados.
- 3) Los algoritmos paralelos, en este orden de ejecución los algoritmos trabajan de manera simultánea.

2.3. Metaheurísticas.

Una metaheurística es una maniobra de alto nivel, la cual usa diversos métodos para la exploración y búsqueda en un espacio, son modelos no deterministas que deben ser completados con datos específicos del problema, las metaheurísticas posibilitan enfrentar problemas con espacios de búsqueda de gran tamaño, En estos métodos se debe establecer que exista una armonía entre intensificación y diversificación. La intensificación es evaluación de soluciones en regiones limitadas y reducidas respecto al espacio de búsqueda enfocada en una zona de soluciones concretas y la diversificación es la evaluación de soluciones en zonas distantes del espacio de búsqueda y la intensificación.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

(García, 2007) define a las metaheurísticas como:

- Estrategias o plantillas generales que guían el proceso de búsqueda.
- El objetivo es una exploración eficiente del espacio de búsqueda para encontrar soluciones (casi) óptimas.
- Las Metaheurísticas son algoritmos no exactos y generalmente son no deterministas.
- Pueden incorporar mecanismos para evitar regiones no prometedoras del espacio de búsqueda.
- El esquema básico de cualquier metaheurística tiene una estructura predefinida.
- Las metaheurísticas hacen uso de conocimiento del problema que se trata de resolver en forma de heurísticos específicos que son controlados por una estrategia de más alto nivel.

Un ejemplo de metaheurística es el algoritmo genético el cual es un método adaptativo, que usa una semejanza directa con el comportamiento natural, este trabaja con una población de individuos, los cuales cada uno representa una solución a un problema. A cada elemento se le asigna una puntuación, relacionando las propiedades dicha solución.

El funcionamiento de los Algoritmos Genéticos en la naturaleza equivaldría a un grado de efectividad de un organismo para competir con otros, cuanto mayor sea la adaptación de un individuo a la situación, mayor será la probabilidad de que este sea seleccionado para reproducirse, y así cruzar su material genético con otro individuo seleccionado de igual manera, así este cruce producirá nuevos individuos descendientes de los anteriores los cuales comparten características de sus padres (Abdelmalik Moujahid, 1859). En cuanto menor sea su adaptación, el individuo tendrá menor probabilidad de reproducirse y por lo tanto que su material genético no se propague en las siguientes generaciones.

En años recientes se han presentado un gran número de algoritmos que no siguen los conceptos puritas de una metaheurísticas tradicional simple, sino que éstos son la combinación de varios algoritmos, los cuales fuera del campo tradicional de las metaheurísticas, son comúnmente llamados metaheurísticas híbridas (Raidl, 2006).

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

2.4. Testores.

Cuando se cuenta con un grupo de datos con demasiadas clases se incrementa la probabilidad de que algoritmos de clasificación encuentren subgrupos de características que no son totalmente correctos (Jensen & Shen, 2008) (Liu & Motoda, 2008). Los métodos de selección pueden acortar el tiempo de entrenamiento de un modelo clasificador (Bermingham et al., 2015), de esta manera se simplifica la interpretación a los usuarios (Casella, Fienberg, & Olkin, 2013), se evita el exceso de dimensión y se generaliza por reducción de varianza.

Los métodos de selección de variables se separan en filter, wrapper e híbridos. Los Métodos wrapper: estos se utilizan para medir o para conducir el proceso de búsqueda de los diferentes subconjuntos de variables, utilizando un algoritmo de agrupamiento y evaluando los resultados mediante criterios específicos. Estos califican las variables en subconjuntos y no solo cada variable por separado.

Los testores típicos se sitúan en estos métodos de selección de variables wrapper, ya que seleccionan los subgrupos de variables y su relación evaluando todas las variables del subgrupo, estos consideran el subconjunto y su interrelación multivariable (MasedaTarin, 2019).

Este concepto aparece a mediados de los años cincuenta, y fue empleado en la detección de fallas de circuitos eléctricos por Cheguis y Yablonskii (Cheguis & Yablonskii, 1955) y después, se utilizó para solucionar problemas de clasificación supervisada y detección de variables en el área de Geología (Alba, Santana, Ochoa, & Lazo, 2000). En los trabajos más recientes se encuentra el trabajo pionero de Dmitriev, Zhuravlev y Krendeleiev en 1966 (Dmitriev, Zhuravlev, & Krendeleiev, 1966).

En este tipo de problemas de aprendizaje supervisado se parte de una matriz inicial de datos dividida en clases y la cual contiene las descripciones de los objetos o fenómenos discretos en N rasgos. En esencia, un testor, es un conjunto de características o rasgos que distinguen elementos (objetos) que pertenecen a clases diferentes.

Un testor típico es el cual, si le elimina un rasgo cualquiera, pierde la propiedad de ser un testor como menciona Pons en 2003 (Santiesteban & Pons, 2003). O de otra forma un testor típico

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

es aquel al que se le va eliminando la redundancia o un testor en su mínima expresión (M. D. Torres, 2010).

El concepto de testor comúnmente asociado al problema de reconocimiento de patrones se encuentra dentro de lo que se conoce como enfoque lógico combinatorio, que permite la determinación de familias y subfamilias de testores: como en el caso de los testores típicos. Para determinar las subfamilias se han desarrollado algunos algoritmos que se dividen en algoritmos de escala interior (Aguila & Ruiz, 1984), (Bravo Martinez, 1983) y los algoritmos de escala exterior, entre los cuales encontramos el algoritmo BT (Shulcloper, Águila, & Bravo, 1985) hay evidencia de que el algoritmo de BT has sido ampliamente utilizado (Sánchez & Lazo, 2002) por ser considerado el paradigma de los algoritmos de escala exterior. Lamentablemente su desempeño computacional no ha sido totalmente conveniente debido a los tiempos de ejecución obtenidos (Sánchez & Lazo, 2002).

El cálculo de un conjunto de testores típicos se considera de complejidad exponencial, en la búsqueda de soluciones para reducir el espacio de búsqueda de la solución y reducción de tiempo menor al requerido en la actualidad, se desarrollan e implementan nuevos algoritmos.

De manera general se dice que los algoritmos de escala interior sirven para encontrar testores típicos analizando la estructura interna de la matriz básica y determinan condiciones para asegurar la condición de testor o testor típico, así mismo los algoritmos de escala exterior encuentran testores típicos por medio de la generación de un grupo potencia del conjunto total de columnas de matriz básica y así el algoritmo es capaz de brincar combinaciones del grupo potencia (Shulcloper, Alba, & Lazo, 1995b).

2.4.1 Método de búsqueda.

Sea M un conjunto finito de objetos sobre los que se definen un conjunto de predicados X_1, \dots, X_r , X_{r+1} . Este último se denomina "predicado fundamental" y su función es determinar la pertenencia de un objeto a una de las clases determinadas. Los restantes predicados se denominarán "rasgos", dado un conjunto de objetos, estos pueden ser puestos en correspondencia con un conjunto de tuplas, así como sus descripciones en valor de los rasgos. Partiendo de las clases K_1, \dots, K_{r+1} si el problema tiene definido el predicado fundamental para algunos de los objetos de cada una de las

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

clases o para todas excepto una, se parte de la descripción estándar de los objetos, en caso de que no se conozcan las l-tuplas de pertenencia de los objetos (Shulcloper et al., 1995b).

Si comenzamos de la comparación de las características de los objetos que pertenecen a la misma clase contra aquellas que pertenecen a las demás clases comparando en pares a la vez, (conjunto de objetos que se definirán como Matriz de Aprendizaje MA), se obtendrá la matriz de diferencias MD. Esta matriz es la que contiene la información que distingue objetos de clases distintas, de ahí proviene su nombre, esta matriz es construida empleando criterios de comparación existentes.

El subconjunto de rasgos o características T de una matriz básica es un testor si al eliminar de la MB todas las características, excepto las pertenecientes a T, no existe ninguna fila de ceros (escaso de información). T es un testor típico, si al quitarle cualquiera de sus características, pierde su condición de testor.

Todos los algoritmos descritos para encontrar todos los testores típicos asociados a un conjunto de datos, tienen complejidad exponencial y dependen del tamaño de la matriz (Sanchez Diaz & Lazo Cortez, 2008).

Usualmente, encontrar todos los testores (típicos) asociados a una matriz de datos, requiere de realizar una búsqueda exhaustiva entre todos los subconjuntos de rasgos que son 2^R , siendo R el conjunto de los rasgos. Con el aumento del número de filas y columnas o del número de clases en la matriz, el costo en tiempo de este proceso puede ser demasiado hasta el punto de ser realmente imposible.

2.4.2 Reglas de comparación.

A continuación, se establecen las reglas de comparación de los objetos de aprendizaje:

Sea $\Omega = \{O_1, O_2, \dots, O_m\}$ un conjunto de m objetos e $I(O_1), I(O_2), \dots, I(O_m)$ sus descripciones en términos de un conjunto de n rasgos o características, $R = \{X_1, X_2, \dots, X_n\}$, donde cada rasgo X_j tiene asociado un conjunto M_j que se designa conjunto de valores admisibles del rasgo X_j , es decir, $I(O_i) = (X_1(O_i), X_2(O_i), \dots, X_n(O_i))$, $X_j(O_i) \in M_j, j = 1, \dots, n, i = 1, \dots, m$ (Santiesteban & Pons, 2003).

CAPÍTULO II. MARCO TEÓRICO

El tipo de una variable depende de la naturaleza de su conjunto de valores admisibles. Los conjuntos M_j pueden ser $\{0,1\}$, el conjunto de los números reales o de los enteros, $\{V_1, \dots, V_k\}$, donde los V_i no son necesariamente números, etc.

A partir de la matriz de aprendizaje (MA) se puede definir otra matriz de diferencias (MD), que explica la igualdad (0) o diferencia (1), rasgo a rasgo cada par de objetos de clases distintas. (Shulcloper, Alba, & Lazo, 1995a).

Se llama criterio de comparación de valores de la variable X_j a una función $C_j: M_j \times M_j \rightarrow L_j$, tal que si C_j es un criterio de disimilaridad se cumple que $C(X(O), X(O)) \min \{y\}$ y si C_j es un criterio de similaridad se cumple que $C(X(O), X(O)) \max \{y\} =$, donde L_j es un conjunto totalmente ordenado, $j = 1, \dots, n$ (Shulcloper, Alba, & Lazo, 1995b).

Se le llama *Matriz de Aprendizaje* (MA) al tuplo $(I(O_1), \alpha(O_1), I(O_2), \alpha(O_2), \dots, I(O_m), \alpha(O_m))$ que incluye las descripciones de los objetos conjuntamente con sus c -tuplos de pertenencia. (Shulcloper et al., 1995b).

El algoritmo BT es un algoritmo de escala exterior, por lo que basa su funcionamiento en los siguientes elementos (Sánchez & Lazo, 2002):

- una caracterización conveniente del concepto de testor y de testor típico.
- un orden definido sobre el conjunto potencia del conjunto de todos los rasgos (R) que describen los objetos en el problema de clasificación que se considera.
- una función de recorrido, que permite encontrar todos los testores típicos sin necesidad de recorrer todo el conjunto potencia.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Para la descripción del algoritmo, se utilizará un n-tuplo booleano $a = (a_1, \dots, a_n)$, tal que si $a_j = 1$, esto indicará que la columna j aparece en el conjunto de columnas a considerar y $a_j = 0$ significa lo contrario. Cada uno de estos n-tuplos representa el llamado vector característico del subconjunto correspondiente de R .

Para encontrar la familia de testores típicos en un problema de clasificación supervisada, utilizaremos la matriz básica (MB) asociada al problema (Shulcloper et al., 1995a)], la cual contiene la información mínima necesaria para encontrar los testores típicos.

Definición 1: El n-tuplo $\alpha = (\alpha_1, \dots, \alpha_n)$ se llamará *lista testor*, si $\forall a = (a_1, \dots, a_n)$ de MB, se cumple que:

$$(\alpha_i \wedge a_i) \neq 0, i = 1, \dots, n \quad \text{Ecuación 1}$$

Donde a es una fila de la matriz básica y $\alpha = (\alpha_1, \dots, \alpha_n)$ representa un testor típico (sabiendo que es lista testor) si para cada j tal que $\alpha_j = 1$, existe alguna fila en MB tal que su única coordenada unitaria coincidente con una coordenada no nula de j -ésima. En este caso α se llamará lista testor típico.

El algoritmo BT considera los n-tuplos característicos, correspondientes a los subconjuntos de rasgos, ordenados ascendentemente, según el orden natural. Desde $(0, 0, \dots, 0, 1)$ hasta $(1, 1, \dots, 1, 1)$.

La función de recorrido, se basa en las siguientes proposiciones, que permiten efectuar saltos (Sánchez & Lazo, 2002):

Proposición 1 (de salto ante testor) sea $\alpha = (\alpha_1, \dots, \alpha_k, \dots, \alpha_n)$ un n-tuplo lista testor, y sea α_k la última coordenada unitaria (el uno más a la derecha) de α . Entonces los siguientes $2^{n-k} - 1$ n-tuplos (en el orden establecido) se corresponden con testores no típicos.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Proposición 2 (de salto ante no testor): Sea $\alpha = (\alpha_1, \dots, \alpha_n)$ un n-tuplo correspondiente con un no testor, y sea $a = (a_1, \dots, a_n)$ una fila de MB tal que $(\alpha_i \wedge a_i) = 0$, donde $i=1, \dots, n$. Sea a_u la última coordenada unitaria de a (el uno más a la derecha), y sea $\alpha' = (\alpha'_1, \dots, \alpha'_n)$.

2.2.3 Ejemplo

El ejemplo presentado a continuación fue tomado textualmente de (Torres Soto, 2010) y está basado en (Santiesteban & Pons, 2003) ya que es una manera sencilla y fácil de entender la obtención de los testores típicos.

Con la siguiente información acerca de los rasgos en términos de los cuales se describen los siguientes objetos (Tabla1):

Tabla 1 Ejemplo de Variables y sus Componentes de Análisis

Rasgo	Tipo	Dominio de definición	Criterio de comparación	Datos adicionales
X1	Continuo	$(-5.5, 5.5)$	error admisible	$\varepsilon=0.5$
X2	Discreto	$[10, 20]$	error admisible	$\varepsilon=2.0$
X3	Booleano	$\{true, false\}$	igualdad estricta	
X4	K-Valente Nominal	$\{azul, blanco, rojo, verde\}$	igualdad estricta	

Una matriz de aprendizaje es la siguiente (Tabla 2):

CAPÍTULO II. MARCO TEÓRICO

Tabla 2 Ejemplo de matriz de Aprendizaje

	X1	X2	X3	X4	α
O1	1.3	10	true	azul	1
O2	-1.3	13	true	rojo	1
O3	2.4	16	false	azul	2
O4	-2.4	19	false	blanco	2
O5	2.6	10	true	verde	3
O6	-2.6	20	false	rojo	3
Criterio Comparación	0.5	2	=	=	

Dada una MA y C_1, \dots, C_n criterios de comparación de cada uno de los rasgos, llamaremos Matriz de Diferencias de MA, y la denotaremos por MD, a la matriz booleana formada por las m' filas (Shulcloper et al., 1995b) y (Shulcloper et al., 1995a)

Donde:

$$S_{ij} = (\delta_1^{ij}, \dots, \delta_n^{ij}), i, j = 1, \dots, m, i \neq j$$

Ecuación 2

Y Donde:

$$\delta_p^{ij} = C_p(X_p(O_i), X_p(O_j)), p = 1, \dots, n$$

Ecuación 3

$$m' = \sum_{i=1}^{c-1} \sum_{t=i+1}^c |K_i| |K_t|$$

Ecuación 4

Sean i_p, i_q filas de MD

Diremos que i_p es subfila de i_q

$$\text{Si } \forall (j) [a_{iqj} = 0 \Rightarrow a_{ipj} = 0] \text{ y además } \exists (j_0) [a_{iqj_0} = 1 \wedge a_{ipj_0} = 0] \quad \text{Ecuación 5}$$

También diremos que i_q es superfila de i_p

Sea i_q una fila de MD. La fila i_q es básica si no existe fila alguna i_p que sea subfila de i_q .

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Dada una matriz de diferencias MD llamaremos Matriz Básica (MB) a la matriz formada exclusivamente por las filas básicas de MD (Shulcloper et al., 1995b).

A continuación, se muestran la Matriz de Diferencia y la Matriz Básica derivadas de la Matriz de Aprendizaje (Tabla 3).

Tabla 3 Ejemplo de Matriz de Diferencias y Matriz Básica

MD	
Filas	Objetos comparados
1110	O1 con O3
1111	O1 con O4
1001	O1 con O5
1111	O1 con O6
1111	O2 con O3
1111	O2 con O4
1101	O2 con O5
1110	O2 con O6
0111	O3 con O5
1101	O3 con O6
1111	O4 con O5
0001	O4 con O6

MB
Filas Básicas
1110
0001

El subconjunto $\tau = \{X_{i1}, \dots, X_{is}\}$ de rasgos de una Matriz de Aprendizaje es un testor si al eliminar de su Matriz Básica todas las columnas, excepto las correspondientes a los elementos de τ , no existe fila alguna completa de ceros.

Para obtener los testores de manera exhaustiva, se debe comparar la matriz básica con cada uno de los miembros elementos del siguiente conjunto (Tabla 4).

CAPÍTULO II. MARCO TEÓRICO

Tabla 4 Ejemplo de Exploración y Conclusión de Candidatos

Elemento	Representación Binaria	Conclusión
1	0001	No es un testor ya que si se proyectan sus características en MB la fila uno de MB seria de ceros
2	0010	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
3	0011	Si es un testor porque sus características si se proyectan en MB no fila de ceros además si se elimina cualquiera de sus características pierde su condición de testor
4	0100	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
5	0101	Si es un testor porque sus características si se proyectan en MB no fila de ceros además si se elimina cualquiera de sus características pierde su condición de testor
6	0110	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
7	0111	Si es un testor porque sus características si se proyectan en MB no fila de ceros además si se elimina cualquiera de sus características pierde su condición de testor, pero no es un testor típico

CAPÍTULO II. MARCO TEÓRICO

8	1000	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
9	1001	Si es un testor porque sus características si se proyectan en MB no fila de ceros además si se elimina cualquiera de sus características pierde su condición de testor
10	1010	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
11	1011	Si es un testor porque sus características si se proyectan en MB no fila de ceros además ya que si se elimina cualquiera de sus características pierde su condición de testor, pero no es un testor típico
12	1100	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
13	1101	Si es un testor porque sus características si se proyectan en MB no fila de ceros además si se elimina cualquiera de sus características pierde su condición de testor, pero no es un testor típico
14	1110	No es un testor ya que si se proyectan sus características en MB la fila dos de MB seria de ceros
15	1111	Si es un testor porque sus características si se proyectan en MB no fila de ceros además si se elimina cualquiera de sus características pierde su condición de testor, pero no es un testor típico

Observe que cualquier conjunto que contenga a un testor típico seguirá siendo testor, pero no típico. Este es definido de fila básica, y la cual no contiene a una menor, es una fila incomparable o mínima. Además, note que las filas marcadas en gris coinciden con los testores típicos de la matriz.

Analice que el conjunto potencia que debe examinarse es de tamaño $2N-1$ siendo N el número total de características que tiene la matriz de entrenamiento.

Estos son los testores encontrados en el ejemplo $\{X_1, X_4\}$, $\{X_2, X_4\}$ y $\{X_3, X_4\}$. Los que pueden ser representados en binario como (Tabla 5):

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Tabla 5 Ejemplo de Testores Encontrados

Testores encontrados	Representación
{X1,X4}	1001
{X2,X4}	0101
{X3,X4}	0011

Para obtener el peso informacional o la importancia de una característica: es posible calcular el factor porcentual que indica cual es la frecuencia de cada variable en el conjunto total de testores. Mediante un factor que nos señala la formación de un subconjunto de características mediante el cual se puede diferenciar a la clase que pertenece un objeto. (M. . Torres, Ponce, Ochoa, Torres, & Díaz, 2009). A continuación, se calcula el peso informacional de cada variable (Tabla 9):

Tabla 6 Ejemplo de Peso Informacional

Variable	Peso informacional
X1	33.33%
X2	33.33%
X3	33.33%
X4	100%

Interpretando la información se puede determinar que la variable 4 es necesaria para la formación de un testor típico ya que su valor o peso informacional es mucho mayor que el de las otras tres variables. Dicho de otra manera, no se puede excluir en ningún momento a la variable 4, si se tiene la información de la variable 1 o 2 o 3, se contará con la información suficiente para diferenciar la clase a la que pertenece cada objeto.

2.5. Redes Neuronales Artificiales

Son modelos computacionales inspirados en funcionamiento neuronal biológico que tienen como aplicación principal el reconocimiento de patrones.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Una red neuronal es un modelo computacional, paralelo, compuesto de unidades procesadoras adaptativas con una alta interconexión entre ellas (Hassoun, 1995).

Internamente, una RNA emula funcionamiento con nodos como neuronas interconectadas unas con otras situadas en diferentes niveles. Que se interconectan unas con otras emulando la liberación de neurotransmisor en espacio situado entre neuronas.

Una RNA es un sistema que permite separar elementos en categorías cuando la diferencia no es muy clara entre ellas; separándolos de manera no lineal.

Es una familia de algoritmos Machine Learning para modelación de comportamientos avanzados.

Una RNA emula la forma de cómo los humanos clasificarnos basado en experiencia

Pasos para que una RNA funcione:

- Crear RNA, establecer una estructura.
- Configurar RNA, dar propiedad de objetos para discriminar.
- Entrenar RNA, dotar de la mayor cantidad posible de ejemplos.
- Clasificar.

En las Redes Neuronales Artificiales (RNA), la unidad análoga a la neurona biológica es el elemento de Procesamiento o Neurona Artificial. Una Neurona Artificial (NA) puede tener varias entradas y combinarlas, generalmente con una suma. La suma de las entradas es modificada por una función de transferencia y el valor de la salida de esta función de transferencia se pasa directamente a la salida del elemento procesador o Neurona Artificial.

Así mismo, la salida de la NA se puede conectar a las entradas de otras NA mediante conexiones ponderadas correspondientes a la eficacia de la sinapsis de las conexiones neuronales.

Una RNA consiste en un conjunto de unidades elementales NA conectadas de una forma concreta. El interés de las RNA no reside solamente en el modelo de la NA sino en las formas en que se conectan estos elementos procesadores.

Generalmente las NA están organizadas en grupos llamados niveles o capas.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Una red característica radica en una secuencia de capas con conexiones entre capas contiguas consecutivas.

En una RNA hay dos capas con conexiones con el exterior.

Una capa de entrada, donde se presentan los datos a la red, y una capa de salida la cual mantiene la respuesta de la red a una entrada.

El resto de las capas son nombradas capas ocultas.

Los elementos de una RNA se muestran en la Figura 2 que está basado en (J. Torres, 2018):

- Entradas.
- Neuronas.
- Capa de entrada.
- Capas profundas u ocultas.
- Capa de salida.
- Salidas.

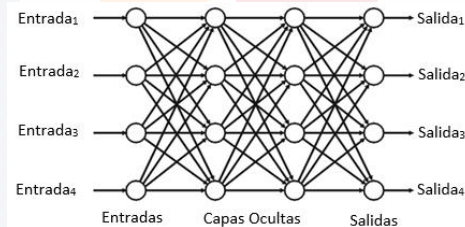


Figura 1 Elementos de RNA Cita

- Neurona Artificial o Unidad básica de procesamiento se muestra en la Figura 3 que está basada en (Ardila, 2017), esta puede ser o no densamente interconectada.
- Entradas: Estas capas reciben la información desde el exterior.
- Pesos: Cada entrada tiene su propio peso relativo o importancia de la entrada.
- Suma ponderada.
- Función de activación.
- Salida.

CAPÍTULO II. MARCO TEÓRICO

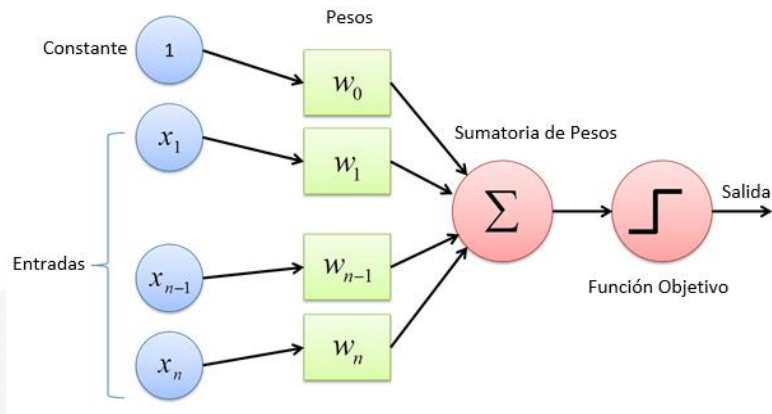


Figura 2 Elementos de Neurona Artificial

2.6 Deep Learning

Las técnicas de machine learning experimentan un incremento en su aplicación tanto en las empresas como en el ámbito académico, ya que éstas son capaces de detectar automáticamente patrones en los datos, ya sea para toma de decisiones o para predicciones.

Existen 4 grupos esenciales de componentes del aprendizaje automático, 1.- Fuentes de Información, 2.- Técnicas y algoritmos de tratamiento de la información, 3.- Capacidad de autoaprendizaje, 4.- Uso de sistemas o software (Management Solutions, 2018).

Dentro de los algoritmos y técnicas de tratamiento de información el aprendizaje automático comprende algoritmos para modelar abstracciones no lineales en forma matricial o tensorial.

Se considera una forma de automatizar el análisis predictivo, estos modelos son conocidos como redes neuronales artificiales Figura 4 la cuál fue tomada de (Barrios Arce, 2020) .

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

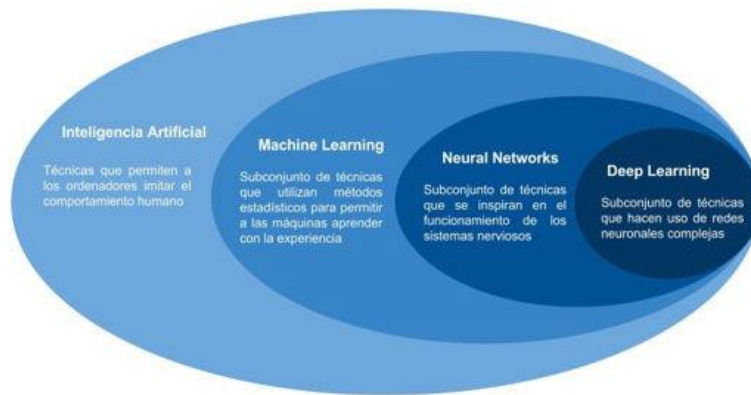


Figura 3 inteligencia artificial

Dentro del área de machine learning o aprendizaje automático, se encuentran las técnicas de Redes neuronales y así mismo estas envuelven a las técnicas de Redes Neuronales profundas, de las cuales se comenta a continuación:

2.6.1 Arquitectura y Características

En otras arquitecturas de computación tradicional las instrucciones se ejecutan sincronizadas y en secuencia con el reloj del sistema. En los sistemas de computación neuronal cada elemento NA sólo puede realizar uno, o como mucho, varios cálculos. La potencia del procesamiento de las RNA se calcula por el número de interconexiones actualizadas por segundo durante el proceso de entrenamiento y aprendizaje.

La arquitectura de las RNA se origina de los sistemas de procesamiento en paralelo, es decir, sistemas en los que diferentes procesadores están interconectados. No obstante, los procesadores son unidades procesadoras simples, diseñadas para la suma de muchas entradas y con un ajuste automático de las conexiones ponderadas para después ser modificadas por la función objetivo.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Para desarrollar una RNA no hay que programar ni el conocimiento ni las reglas del procesamiento del conocimiento. Una RNA aprende las reglas del procesamiento del conocimiento a través del ajuste de conexiones ponderadas entre las neuronas de las distintas capas de la red.

Una característica importante de las RNA es la forma en que se guarda la información. La memoria o el conocimiento de estas redes está distribuida a lo largo de todas las conexiones ponderadas de la red.

Una propiedad de la naturaleza de la memoria de las RNA es llamada Generalización, posibilita que la red reaccione adecuadamente cuando se le presenta una entrada con información incompleta o con ruido.

La tolerancia a la falta es otra característica de las RNA es la tolerancia a la falta. Esta tolerancia a la falta se refiere al hecho de que en una RNA resultaran erróneos o destruidos algunas NA, o se alteraran las conexiones entre estas, el comportamiento de la red sería mínimamente afectado o modificado. Ya que el comportamiento varía, pero él no deja de funcionar.

Esta característica se debe a que las RNA tienen la información distribuida a lo largo de toda la red y no está contenida en un único lugar.

Hay dos fases en la modelización con redes neuronales como lo menciona (Marín Diazaraque, 2003):

Fase de entrenamiento: se usa un conjunto de datos o patrones de entrenamiento para determinar los pesos (parámetros) que definen el modelo de red neuronal. Se calculan de manera iterativa, de acuerdo con los valores de los valores de entrenamiento, con el objeto de minimizar el error cometido entre la salida obtenida por la red neuronal y la salida deseada.

Fase de Prueba: en la fase anterior, el modelo puede que se ajuste demasiado a las particularidades presentes en los patrones de entrenamiento, perdiendo su habilidad de generalizar su aprendizaje a casos nuevos (sobre ajuste).

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Para evitar el problema del sobre ajuste, es aconsejable utilizar un segundo grupo de datos diferentes a los de entrenamiento, el grupo de validación, que permita controlar el proceso de aprendizaje.

2.6.2 Entrenamiento en Redes Neuronales

La capacidad de aprendizaje es una de las características de las Redes Neuronales Artificiales (RNA), El entrenamiento de las RNA tiene parecido con el progreso intelectual de los seres humanos. Incluso cuando se crea que se tiene el dominio y entendimiento del proceso de aprendizaje es conveniente ser moderados, ya que el aprendizaje de las RNA es acotado.

La meta del entrenamiento de una RNA es lograr que una aplicación, para un grupo de entradas produzca el conjunto de salidas deseadas o mínimamente consistentes.

El proceso de entrenamiento consta en aplicar secuencialmente diferentes conjuntos o vectores de entrada para que se ajusten los pesos de las interconexiones según una técnica predeterminada. En el proceso de entrenamiento los pesos tienden gradualmente hacia los valores que hacen que cada entrada produzca el vector de salida deseado.

Los algoritmos de entrenamiento o los procedimientos de ajuste de los valores de las conexiones de las RNA se pueden clasificar en dos grupos: Supervisado y No Supervisado como lo menciona a continuación (Basogain, 2005) .

Entrenamiento Supervisado: estos algoritmos requieren el emparejamiento de cada vector de entrada con su correspondiente vector de salida. El entrenamiento consiste en presentar un vector de entrada a la red, calcular la salida de la red, compararla con la salida deseada, y el error o diferencia resultante se utiliza para realimentar la red y cambiar los pesos de acuerdo con un algoritmo que tiende a minimizar el error.

Las parejas de vectores del conjunto de entrenamiento se aplican secuencialmente y de forma cíclica. Se calcula el error y el ajuste de los pesos por cada pareja hasta que el error para el conjunto de entrenamiento entero sea un valor pequeño y aceptable.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Entrenamiento No Supervisado: los sistemas neuronales con entrenamiento supervisado han tenido éxito en muchas aplicaciones y sin embargo tienen muchas críticas debido a que desde el punto de vista biológico no son muy lógicos. Resulta difícil creer que existe un mecanismo en el cerebro que compare las salidas deseadas con las salidas reales. En el caso de que exista, ¿de dónde provienen las salidas deseadas?

Los sistemas no supervisados son modelos de aprendizaje basados en los sistemas biológicos desarrollados por (Teuvo Kohonen, 1988). Estos sistemas de aprendizaje no supervisado no requieren de un vector de salidas deseadas y por tanto no se realizan comparaciones entre las salidas reales y salidas esperadas.

Entre los algoritmos de entrenamiento que existen hoy en día, existe uno que cabe la pena mencionar el cual fue propuesto por Hebb en el cual se suma o incrementa el peso de la conexión cuando dos o más neuronas conectadas son activadas (HEBB, 1992).

Perceptrón

El Perceptrón tiene dos capas de unidades procesadoras (PE) y solo una de ellas tiene la habilidad de adaptar o modificar los pesos de las conexiones. La arquitectura del Perceptrón admite capas adicionales (Barrios Arce, 2020).

2.6.3 Función de Activación.

Una neurona biológica puede estar activa (excitada) o inactiva (no excitada); es decir, que tiene un “estado de activación”. Las neuronas artificiales también tienen diferentes estados de activación; algunas de ellas solamente dos, al igual que las biológicas, pero otras pueden tomar cualquier valor dentro de un conjunto determinado.

La función activación calcula el estado de actividad de una neurona; transformando la entrada global (menos el umbral) en un valor (estado) de activación, cuyo rango normalmente va de (0 a 1) o de (-1 a 1). Esto es así, porque una neurona puede estar totalmente inactiva (0 o -1) o activa (1) (Matich, 2001).

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Entre las funciones de activación más comunes clasificadas por (Match, 2001) se encuentran las siguientes:

1. Función lineal en la Figura 5 tomada de (Match, 2001).

$$f(x) = \begin{cases} -1 & x \leq -1/a \\ a * x & -1/a < x < 1/a \\ 1 & x \geq 1/a \end{cases}$$

Ecuación 6

con $x = gin_i - \Theta_i$, y $a > 0$.

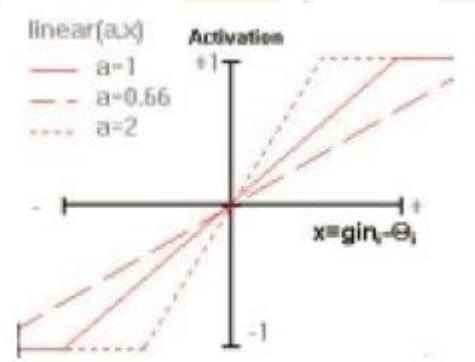


Figura 4 Función lineal

Por encima o por debajo de esta zona se fija la salida en 1 o -1, respectivamente.

Cuando $a = 1$, la salida es igual a la entrada.

2. Función Sigmoide en la Figura 6 tomada de (Match, 2001).

$$f(x) = \frac{1}{1 + e^{-gx}}, \text{ con } x = gin_i - \Theta_i.$$

Ecuación 7

CAPÍTULO II. MARCO TEÓRICO

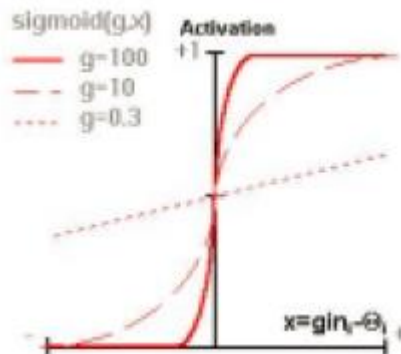


Figura 5 Función Sigmoide

En esta función al cambiar el valor de g se ve modificada la pendiente de la función de activación y los valores de salida comprenden de un rango que va de 0 a 1.

3. Función Tangente Hiperbólica en la Figura 7 tomada de (Matich, 2001).

$$f(x) = \frac{e^{gx} - e^{-gx}}{e^{gx} + e^{-gx}}, \text{ con } x = gin_i - \Theta_i.$$

Ecuación 8

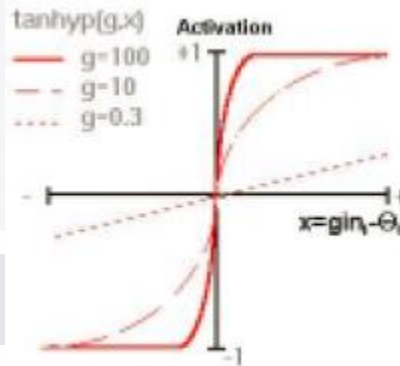


Figura 6 Función Tangente Hiperbólica

Mediante esta función los valores de salida están comprendidos dentro de un rango que va de -1 a 1. Y al igual que la sigmoide al cambiar el valor de g se ve modificada la pendiente de la función de activación.

CAPÍTULO II. MARCO TEÓRICO

2.6.4 Tipos de Redes Neuronales más Comunes.

DNN (Deep Neuronal Networks).

Son utilizadas para procesamiento de texto, imágenes y datos numéricos, tienen una capa de entrada 2 o más capas ocultas y una capa de salida, su restricción es que debido a la gran cantidad de conexiones entre las capas cuando se tienen muchas entradas, se vuelven computacionalmente más pesadas Figura 8 basada en (J. Torres, 2018)

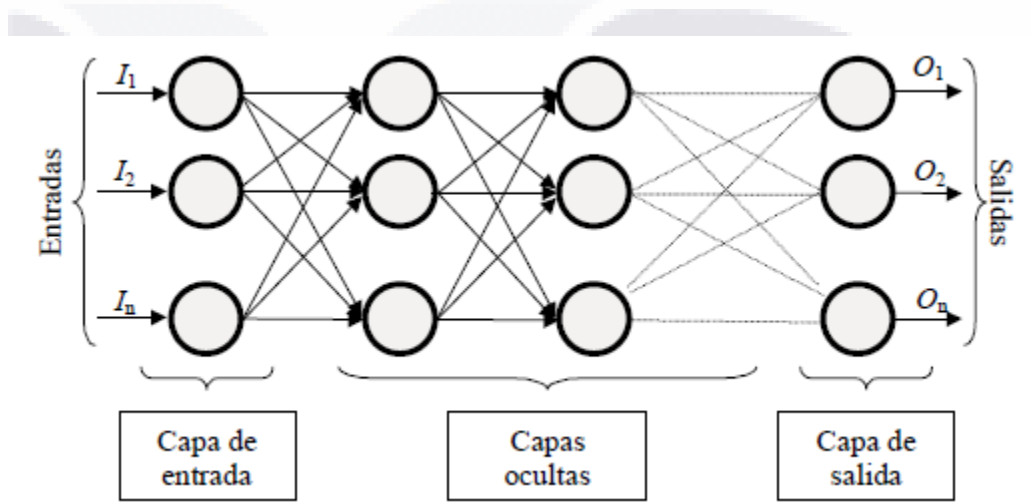


Figura 7 DNN

CNN (Convolutional Neuronal Networks).

Estas redes implementan convoluciones, submuestreo de datos max pooling para disminuir la dimensionalidad del problema, aporta variación a traslación, en cada capa se reduce el tamaño de la imagen conservando elementos importantes, se puede perder información muy útil, estas son utilizadas para procesamiento de imágenes y texto(Lecun, Bottou, Bengio, & Ha, 1998). en Figura 9 tomada de (Gonzalez Gonzalez, 2022)

CAPÍTULO II. MARCO TEÓRICO

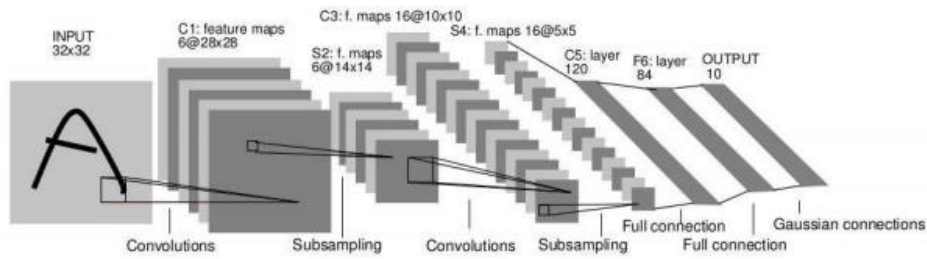


Figura 8 CNN

RNN (Recurrent Neuronal Networks).

Conocidas como redes realimentadas, son redes que permiten auto conexiones o lazos de realimentación en las neuronas, conexiones hacia atrás en las capas, además tienen la capacidad de modelizar comportamientos dinámicos, como se muestra en la figura 9 de (Serrano, Soria, & Martin, 2010). Otro beneficio es que se pueden obtener resultados comparables con redes de mayor tamaño sin realimentación, Son estructuras estáticas sin habilidad para capturar características temporales de los patrones (Serrano, Soria, & Martin, 2010), son comúnmente usadas para tipos de datos secuenciales Figura 10.

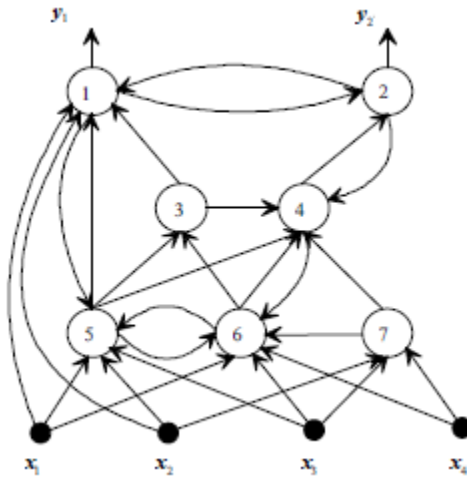


Figura 9 RNN

2.7. Clasificadores Bayesianos.

Clasificación consiste en asignar un objeto a una clase. Por ejemplo, se puede asignar palabras a categorías gramaticales: sustantivo, verbo, adjetivo, etc. también se puede clasificar una imagen

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

como escenario, retrato, etc. El clasificar lo que percibimos a través de nuestros sentidos es algo natural como ser humano, inicialmente esto nos permite abstraer la información, trasladándola a una representación más adecuada para la toma de decisiones.

Los clasificadores bayesianos o redes bayesianas, así como las redes neuronales artificiales y los árboles de decisión, han sido las tres técnicas más usadas en aprendizaje automático en estos últimos años.

Es un clasificador estadístico clásico, éstos, además de ofrecer el análisis cualitativo de los atributos y valores que pueden intervenir en el problema, integran también la importancia cuantitativa de los atributos.

El aprendizaje basado en clasificadores bayesianos es especialmente adecuado en ciertas tareas como la clasificación de textos.

Desde un enfoque bayesiano, el problema de clasificación supervisada consiste en asignar a un objeto descrito por un conjunto de atributos o características, X_1, X_2, \dots, X_n , a una de m clases posibles, c_1, c_2, \dots, c_m , tal que la probabilidad de la clase dados los atributos se maximiza como lo explica (Sucar & Titular, 2013) mediante 5 ecuaciones:

$$\text{ArgC}[\text{MaxP}(C \text{ j } X_1, X_2, \dots, X_n)] \quad \text{Ecuación 9}$$

Si denotamos el conjunto de atributos como: $X = \{X_1, X_2, \dots, X_n\}$, la ecuación 1 se puede escribir como: $\text{ArgC}[\text{MaxP}(C \text{ j } X)]$. La formulación del clasificador bayesiano se base en utilizar la regla de Bayes para calcular la probabilidad posterior de la clase dados los atributos:

$$P(C \text{ j } X_1, X_2, \dots, X_n) = P(C)P(X_1, X_2, \dots, X_n \text{ j } C) = P(X_1, X_2, \dots, X_n) \quad \text{Ecuación 10}$$

Que se puede escribir de forma más compacta como:

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

$$P(C \text{ j } X) = P(C)P(X \text{ j } C)=P(X) \quad \text{Ecuación 11}$$

Así que el problema de clasificación basado en la ecuación 3, se puede expresar como:

$$\text{ArgC}[\text{Max}[P(C \text{ j } X) = P(C)P(X \text{ j } C)=P(X)]] \quad \text{Ecuación 12}$$

El denominador, $P(X)$, no varía para las diferentes clases, por lo que se puede considerar como una constante si lo que interesa es maximizar la probabilidad de la clase:

$$\text{ArgC}[\text{Max}[P(C \text{ j } X) = _P(C)P(X \text{ j } C)]] \quad \text{Ecuación 13}$$

Basado en la ecuación 5, para resolver un problema de clasificación bajo el enfoque bayesiano, se requiere la probabilidad a priori de cada clase, $P(C)$, y la probabilidad de los atributos dada a clase, $P(X \text{ j } C)$, conocida como verosimilitud; para obtener la probabilidad posterior $P(C \text{ j } A)$. Entonces, para aprender este clasificador de un conjunto de datos, se requiere estimar estas probabilidades, a priori y verosimilitud, a partir de los datos, conocidos como los parámetros del clasificador.

La aplicación directa de la ecuación 5, resulta en un sistema muy complejo al implementarlo en una computadora, ya que el término $P(X_1, X_2, \dots, X_n \text{ j } C)$, incrementa exponencialmente de tamaño en función del número de atributos; resultando en un requerimiento muy alto de memoria para almacenarlo en una computadora, y también el número de operaciones para calcular la probabilidad crece significativamente. Una alternativa es considerar ciertas relaciones de dependencia mediante lo que se conoce como el clasificador bayesiano simple.

La idea de usar el teorema de Bayes en cualquier problema de aprendizaje automático (en especial los de clasificación) es que podemos estimar las probabilidades a posteriori de cualquier

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

hipótesis consistente con el conjunto de datos de entrenamiento para así escoger la hipótesis más probable. (Malagón, 2003).

2.7.1 Naïve Bayes.

Es el Clasificador basado en probabilidad más simple para aprendizaje supervisado.

Dado un ejemplo x representado por k valores, el clasificador naïve Bayes busca encontrar la hipótesis más probable que describa a ese ejemplo.

Tiene una estructura de pocas clases, muchos atributos, puede ser polinomial, maneja variables discretas o continuas.

Está basado en el Teorema de Bayes, permite revisar probabilidades previamente calculadas cuando se posee nueva información, posibilita inferir la probabilidad de un suceso con base en el conocimiento que se tiene de sucesos relacionados.

Tiene como ventajas, bajo tiempo de clasificación, bajo tiempo de aprendizaje, bajos requerimientos de memoria, sencillez, buenos resultados, buena precisión.

Clasificador Naïve Bayes

$$P(Y|X) = \frac{P(X|Y) P(Y)}{P(X)}$$

Ecuación 14

X y Y : son sucesos relacionados

$P(X)$ y $P(Y)$: probabilidades de X y Y

$P(X|Y)$ y $P(Y|X)$: probabilidades de que ocurra uno y en otro.

2.8. Diabetes Mellitus.

La diabetes Mellitus, o simplemente la diabetes, es una enfermedad crónica que aparece cuando el páncreas no puede producir insulina o cuando el cuerpo no puede hacer un buen uso de la insulina

CAPÍTULO II. MARCO TEÓRICO

que produce. La insulina es una hormona producida por el páncreas que actúa como una llave que permite que la glucosa de los alimentos que ingerimos pase de la sangre a las células del cuerpo para producir energía. Todos los alimentos ricos en hidratos de carbono se descomponen en glucosa en la sangre. La insulina ayuda a la glucosa a entrar en las células. (INEGI, 2011).

La incapacidad de producir insulina o de utilizarla de manera eficaz conduce a niveles elevados de glucosa en sangre (conocida como hiperglucemia). Los altos niveles de glucosa durante un período de tiempo prolongado se asocian con daños corporales y fallos en varios órganos y tejidos. (INEGI, 2011).

La diabetes se está convirtiendo rápidamente en la epidemia del siglo XXI y en un reto de salud global. Estimaciones de la Organización Mundial de la Salud indican que, a nivel mundial, de 1995 a la fecha casi se ha triplicado el número de personas que viven con diabetes, con cifra actual estimada en más de 347 millones de personas con diabetes.^{1, 2} De acuerdo con la Federación Internacional de Diabetes, China, India, Estados Unidos, Brasil, Rusia y México, son en ese orden los países con mayor número de diabéticos (Herrera, 2009).

La Organización Mundial de la Salud posiciona a la diabetes mellitus es a nivel mundial como la Octava causa de muerte del año 2000 al 2012, el INEGI sitúa a la diabetes mellitus como una de las 3 causas de muerte desde 2006 a nivel nacional, el Gobierno del Estado de Aguascalientes la posiciona como la tercera causa en mortalidad general y en primer lugar como causa única de muerte (INEGI, 2013).

En 2016 se realizó un estudio acerca niveles de diabetes en todo el mundo, dirigido por científicos del Imperial College de Londres, de la escuela de salud pública Harvard T.H. Chan, de la OMS y otros 500 investigadores de todo el mundo este estudio menciona que en 2016 el costo anual del tratamiento de diabetes fue de más de 825 billones de dólares nivel mundial, 170 billones en china, 105 billones en Estados Unidos de América y 73 billones de dólares en la India.

Este trabajo se enfocó principalmente a pacientes diagnosticados con diabetes tipo 2, en los cuales los factores de riesgo según El centro médico de La universidad de Maryland (MU, 2013) publica que son:

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

- Edad superior a 45 años
- Diabetes durante un embarazo previo
- Peso corporal excesivo (especialmente alrededor de la cintura)
- Antecedentes familiares de diabetes
- Dar a luz un bebé que pese más de 4 kg (9 libras)
- Colesterol HDL de menos de 35 mg/dL
- Niveles sanguíneos altos de triglicéridos, un tipo de molécula de grasa (250 mg/dL o más)
- Hipertensión arterial (superior o igual a 140/90 mmHg)
- Trastorno en la tolerancia a la glucosa
- Bajo nivel de actividad (hacer ejercicio menos de tres veces a la semana)
- Síndrome metabólico
- Poliquistosis ovárica
- Una afección llamada acantosis pigmentaria, la cual provoca oscurecimiento y engrosamiento de la piel alrededor del cuello o las axilas.

La federación Internacional de diabetes (FID, 2015) por su parte menciona que los factores de riesgo asociados a la diabetes mellitus tipo 2 son:

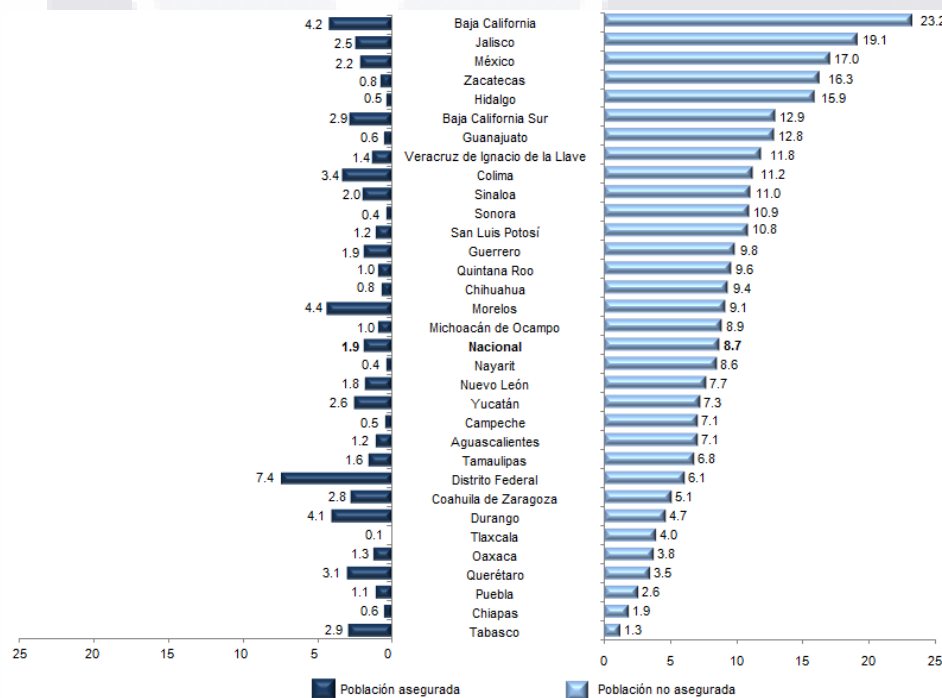
- Antecedentes familiares de diabetes
- Sobrepeso
- Dieta poco sana
- Inactividad física
- Edad avanzada
- Presión arterial alta
- Origen étnico
- Tolerancia anormal a la glucosa (TAG)*
- Antecedentes de diabetes gestacional
- Mala nutrición durante el embarazo

Dentro del artículo del día internacional de la Diabetes emitido por el INEGI se describe a la diabetes mellitus como una enfermedad crónico degenerativa que se presenta cuando el páncreas

CAPÍTULO II. MARCO TEÓRICO

no produce insulina, o bien, la que se produce no es utilizada de manera eficiente por el organismo; ésta es la hormona responsable de que la glucosa de los alimentos sea absorbida por las células y dotar de energía al organismo. Además, menciona que el 1.5% de población de 20 años y más que acudió a servicios de medicina preventiva para detección de diabetes, dio un resultado positivo

A continuación, se muestra una gráfica de (INEGI, 2014) en la figura 11 que presenta el porcentaje de población de 20 años y más con exámenes positivos de diabetes mellitus, por entidad federativa según condición de aseguramiento.



Fuente: SSA, SINAIS. (2013). *Boletín de Información Estadística*. Vol. III. Servicios Otorgados y Programas Sustantivos No. 31. Proceso INEGI.

Figura 10 Incidencia en México de Diabetes Mellitus Grupo de Edad y Sexo

Otra tabla interesante de (INEGI, 2014) presentada en Tabla 7 es la de Incidencia de diabetes mellitus por grupo de edad según sexo del año 2011 en la cual los datos se presentan por cada 100 mil habitantes de cada grupo de edad y sexo.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Tabla 7 Incidencia en México de Diabetes Mellitus Grupo de Edad y Sexo

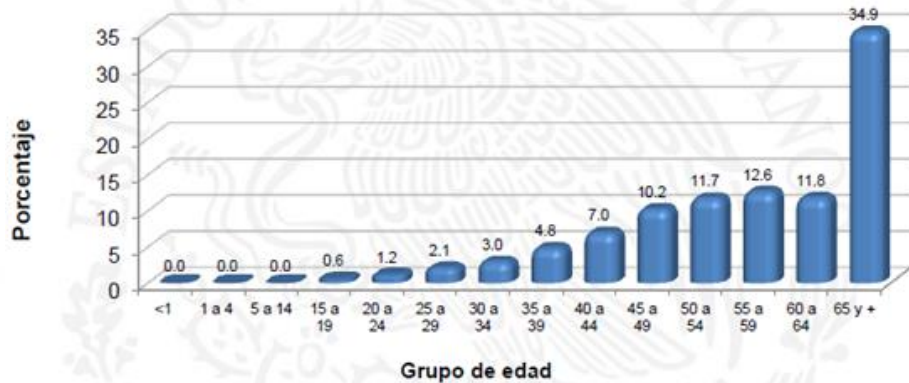
GRUPO DE EDAD	TOTAL	HOMBRES	MUJERES
Menores de 10 años	3.83	3.5	4.13
10 a 14	10.85	8.85	12.93
15 a 19	17.73	14.14	21.34
20 a 24	45.59	34.66	58.02
25 a 44	280.47	240.69	316.67
45 a 49	977.69	828.15	113.28
50 a 59	1363.83	1663.67	1545.06
60 a 64	1787.60	1636.57	1924.23
65 y más	1249.29	1160.42	1325.11

Revisando las investigaciones anuales respecto a diabetes mellitus existe el observatorio nacional de lesiones, el cual lleva el registro de los decesos a nivel nacional, al revisar el reporte del municipio de Aguascalientes se encontró la Diabetes Mellitus y sus complicaciones como primera causa de muerte del 2006 hasta el 2010 (INEGI, 2014).

La Subsecretaría de Prevención y Promoción de la Salud, en su boletín Epidemiológico de Diabetes Mellitus tipo 2 menciona que la IFD (federación internacional de Diabetes Mellitus) reportó en el año 2012 un gasto de 471 miles de millones de dólares.

A continuación, en la figura 12 de (Ejecutivo & Estado, 2011) se presenta la comparación del porcentaje de casos hospitalizados con Diabetes Mellitus por grupo de edad de enero a marzo 2013.

CAPÍTULO II. MARCO TEÓRICO



Fuente: SVEHDMDGAE/Secretaría de Salud/México, enero-marzo 2013.

Figura 11 Casos de Hospitalización por Diabetes Mellitus por Grupo de Edad

Para clasificar el diagnóstico de los pacientes existe la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud, Décima Revisión (**CIE-10**) (ORGANIZACIÓN PANAMERICANA DE LA SALUD, 2015), la cual fue respaldada por la Cuadragésima Tercera Asamblea Mundial de la Salud en mayo de 1990 y se empezó a usar en los Estados Miembros de la Organización Mundial de la Salud (OMS) a partir de 1994.

La Organización Panamericana de la Salud pública, (ORGANIZACIÓN PANAMERICANA DE LA SALUD, 2015) establece que el propósito del CIE es permitir el registro sistemático, el análisis, la interpretación y la comparación de los datos de mortalidad y morbilidad recolectados en diferentes países o áreas, y en diferentes momentos. La clasificación permite la conversión de los términos diagnósticos y de otros problemas de salud, de palabras a códigos alfanuméricos de 4 dígitos que facilitan su almacenamiento y posterior recuperación para el análisis de la información.

Al realizar una revisión en la clasificación se encontró que los diagnósticos propios de diabetes mellitus tipo 2 abarcan el rango de clasificación de E10x hasta E15x como se muestra en su descripción en el Anexo A.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

2.9. K-Means.

Los algoritmos de segmentación son parte del grupo de métodos de minería de datos determinados como no supervisados. El objetivo de la agrupación no es clasificar, predecir o determinar una variable, sino comprender la estructura macroscópica y las relaciones que existen entre los objetos, considerando las formas en que son similares y diferentes. En otras palabras, se centra en dividir el conjunto de datos en subgrupos homogéneos.

El Clustering tiene varias características, entre las que destacan las siguientes:

- Escalabilidad: estos algoritmos funcionan tanto en situaciones en las que hay pocos datos, como en otras con muchos.
- Clústeres de forma arbitraria: dependiendo de la función aplicada para la diferenciación, es la forma que puede adoptar el clúster. Los que se basan en distancias numéricas tienden a encontrar clústeres esféricos.
- Capacidad de manejar diferentes tipos de atributos: numéricos (los más comunes), binarios, nominales, ordinales, etc.
- Capacidad de añadir restricciones: permiten incorporar diversas restricciones para obtener mejores resultados, ya sea limitando la agrupación, seleccionando diversas formas de medición o incorporando nuevas variables.
- Manejo del ruido: muchos de estos algoritmos son sensibles a los datos erróneos. Pueden trabajar eficazmente con datos de alta dimensionalidad.
- Son independientes del orden de los datos.
- Los clústeres son interpretables y utilizables.

Existen varios métodos de agrupación:

- Jerárquico: los datos se agrupan en forma de árbol (top-down o bottom-up).
- No jerárquico: se generan particiones a un solo nivel (K-means).
- Paramétrico: se supone que las densidades condicionales de los grupos tienen una determinada forma paramétrica conocida y se reduce a estimar los parámetros
- No paramétrico: no se hace ninguna suposición sobre la forma de agrupación de los objetos.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Este algoritmo K-means, creado por MacQueen en 1967, es el algoritmo de clusterización más famoso y empleado, ya que a pesar de su simpleza es eficaz. La clusterización permite el reconocimiento de tipologías o grupos en los cuales los elementos comparten gran similitud entre ellos y muchas diferencias con los elementos de los de otros grupos (Salcedo López, 2015).

K-means lleva a cabo un procedimiento de clasificación de un conjunto de elementos en un determinado número K de clústeres o grupos, anticipadamente definido. Se le da el nombre de K-means o K –medias y que representa cada uno de los grupos por la media de los puntos de los elementos, o dicho de otra manera por sus centroides. La interpretación por medio de centroides tiene la utilidad de que es un concepto estadístico y grafico reconocido. Un clúster por lo tanto es determinado por su centroide el cual se encuentra justo en el centro de los elementos que componen el clúster (Mamani, 2015).

Algunos autores como Cifuentes Ramos (Cifuentes Ramos, 2016) dividen la técnica de K-means en las siguientes 4 etapas:

1. Se eligen al azar los K elementos que constituyen los K clúster o grupos iniciales. El valor inicial del centro para cada clúster k es X_i , y hasta este punto este sería el único objeto perteneciente al clúster o grupo.
2. Se debe determinar los elementos al clúster, donde para cada elemento se determina al objeto más cercano al objeto, tomando en cuenta una medida de distancia.
3. Una vez que todos los elementos son situados, se calculan los centroides del K clústeres. Estos nuevos clústeres representan la media del total de elementos designados al clúster.
4. Volver a hacer las etapas 2 y 3 hasta que no se tenga diferencias en reasignaciones.

El algoritmo siempre finaliza, pero eso no asegura el llegar la solución óptima. Se debe tomar en cuenta que para tener mejores resultados se debe ejecutar el algoritmo varias veces para minimizar la sensibilidad a la elección aleatoria de K centroides iniciales.

2.10. CIE 10.

Es la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud Décima Revisión, ésta tiene el propósito de permitir un registro sistemático, interpretación, análisis

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

y comparación de los datos de la mortalidad y morbilidad en diferentes países y épocas. Permite clasificar enfermedades y problemas relacionados con la salud, así como diagnósticos, razones de admisión, afecciones tratadas, motivo de la consulta e incluso las causas de defunción.(Vélez, 2016)

"La CIE 10 Utiliza códigos alfanuméricos los cuales convierten elementos o términos de diagnósticos médicos, lo cual permite su fácil almacenamiento y posterior recuperación para el análisis de información". Esto permite el análisis de salud en grupos de poblaciones, así como el conocimiento de incidencia y persistencia de patologías.

Esto beneficia como una herramienta para toma de decisiones, una forma de exploración de diferentes soluciones a patologías en el momento, además se obtienen estadísticas más específicas, así como un efectivo análisis de mortalidad y morbilidad.

Esta está integrada por 3 volúmenes: vol. 1 Clasificaciones, vol. 2 Instrucciones y orientación y vol. 3 índice alfabético.

La organización del formato de esta clasificación está determinada por 4 dígitos, el primer dígito determina el capítulo por (E es el capítulo 4 Enfermedades endocrinas, nutricionales y metabólicas), el conjunto de los primeros 3 dígitos determina el grupo (Diabetes Mellitus E10 - E14), el tercer dígito establece la categoría (E11 Diabetes Mellitus no Insulinodependiente) , y el cuarto dígito es la subcategoría o complicación (E114 Diabetes Mellitus no Insulinodependiente con complicaciones renales). (Panamericana, 2003)

En el CIE10 se extiende la información referente a diabetes de la siguiente manera:

Diabetes Mellitus (E10-E14)

Las siguientes subdivisiones de cuarto carácter son para ser usadas con las categorías E10-E14:

- 0. Con Coma
 - Diabético
 - Con o sin cetoacidosis.
 - Hipesomolar
 - Hipoglucémico

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

- Hiperglicémico
1. Con Cetoacidosis
 - Acidosis
 - Cetoacidosis
 2. Con complicaciones renales
 - Glomerulonefrosis intracapilar(N083*)
 - Nefropatía diabética(N083*)

Síndrome de Kimmelstiel-Wilson(N083*)

3. Con complicaciones oftálmicas
 - Catarata(H280*) Diabética
 - Retinopatía(H360*) Diabética
4. Con complicaciones neurológicas.
 - Aminotrofia(G730*)
 - Mononeuropatía autonómica(G990*)
 - Polineuropatía(G632*)
 - autonómica(G990*)
5. Con complicaciones circulatorias periféricas
 - Angiopatía periférica(I792*)
 - Gangrena
 - Ulcera
6. Con otras complicaciones especificadas
 - Artropatía diabética(M142.*)
 - Neuropática(M146*)
7. Con complicaciones múltiples
8. Con complicaciones no especificadas
9. Sin mención de complicación

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

E10 Diabetes mellitus Insulinodependiente

Incluye: diabetes (mellitus) (con):

- Juvenil
- Lábil
- Propensión a cetosis

Tipo 1

Excluye: alteración de la tolerancia a la glucosa (R730)

Diabetes mellitus (en):

- Asociada a desnutrición(E12-)
- Embarazo, parto y puerperio(O24-)
- Neonatal(P702)

Glucosuria:

- SAI(R81-)
- Renal(E748)
- Hipoinsulemia postquirúrgica(E891)

E11 Diabetes mellitus no Insulinodependiente

Incluye: diabetes (mellitus) (sin obesidad) (con obesidad):

- De comienzo en el adulto
- De comienzo en la madurez del adulto
- Estable
- No cetósica
- Tipo II

Diabetes no Insulinodependiente juvenil

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Excluye: alteración de la tolerancia a la glucosa (R730)

Diabetes mellitus (en):

- Desnutrición(E12-)
- Embarazo, parto y puerperio(O24-)
- Neonatal(P702)

Glucosuria:

- SAI(R81-)
- Renal(E748)
- Hipoinsulemia postquirúrgica(E891)

E12 Diabetes mellitus asociada con desnutrición

Incluye: desnutrición relacionada con diabetes mellitus:

- Insulinodependiente
- No Insulinodependiente

Excluye: alteración de la tolerancia a la glucosa (R730)

Diabetes mellitus (en):

- Desnutrición(E12-)
- Embarazo, parto y puerperio(O24-)
- Neonatal(P702)

Glucosuria:

- SAI(R81-)
- Renal(E748)
- Hipoinsulemia postquirúrgica(E891)

E13 Otras diabetes mellitus especificadas

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

Excluye: alteración de la tolerancia a la glucosa (R730)

Diabetes mellitus (en):

- Asociada con desnutrición(E12-)
- Embarazo, parto y puerperio(O24-)
- Neonatal(P702)
- No Insulinodependiente (E11-)

Glucosuria:

- SAI(R81-)
- Renal(E748)
- Hipoinsulemia postquirúrgica(E891)

E14 Diabetes mellitus, no especificada

Incluye: Diabetes SAI

Excluye: alteración de la tolerancia a la glucosa (R730)

Diabetes mellitus (en):

- Asociada con desnutrición(E12-)
- Embarazo, parto y puerperio(O24-)
- Neonatal(P702)
- No Insulinodependiente (E11-)

Glucosuria:

- SAI(R81-)
- Renal(E748)
- Hipoinsulemia postquirúrgica(E891).

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

2.11. Métodos de decisión multicriterio.

Existen varias opciones de métodos de toma de decisiones, por lo que es esencial conocer y comprender las diferencias y aportaciones de cada una de las opciones disponibles. Estos métodos utilizan técnicas numéricas que ayudan a los responsables de la toma de decisiones a seleccionar entre un conjunto discreto de decisiones alternativas.

Estos métodos se basan en el impacto de las alternativas. Comparar estos métodos de decisión para elegir el mejor suele ser una tarea complicada; estos modelos se han utilizado ampliamente a pesar de las críticas que existen en torno a ellos.

Entre los modelos más utilizados, podemos encontrar uno de los más antiguos, el modelo de suma ponderada (WSM), Una modificación de este modelo sus puntos débiles son el modelo de producto ponderado (WPM), Otro modelo es el propuesto por Saaty el proceso jerárquico analítico (AHP), Otros métodos muy utilizados son los métodos ELECTRE y TOPSIS.

Para utilizar un método de toma de decisiones que implique un análisis numérico detallado de las opciones, hay que seguir tres pasos:

1. Se determinan los criterios y las opciones.
2. Se asignan medidas numéricas a la importancia relativa de los criterios y a las consecuencias de las opciones sobre estos criterios.
3. Se procesan los valores numéricos y se determina la clasificación de cada opción.

2.12. Técnica de Preferencia de Orden por Similitud a la Solución Ideal.

Dentro de la diversidad de técnicas de apoyo a la selección en condiciones multicriterio se encuentra TOPSIS la Técnica de Preferencia de Orden por Similitud a la Solución Ideal. (Triantaphyllou, 2000). El concepto principal de TOPSIS es que la mejor solución es la que tiene la menor distancia a la solución ideal y la que tiene la mayor distancia a la solución anti-ideal (Olson, 2004) TOPSIS fue desarrollado por Hwang y Yoon en 1981 (Ishizaka & Nemery, 2013).

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO II. MARCO TEÓRICO

La técnica TOPSIS se ha aplicado a una diversa gama de problemas de decisión, por ejemplo, en el apoyo a la inversión financiera, en los sistemas avanzados de fabricación, en la aplicación de selección de procesos robóticos, en los enfoques de redes neuronales, en las extensiones de conjuntos difusos, también se ha utilizado para comparar el rendimiento de las empresas y el rendimiento de los ratios financieros en industrias específicas.(Olson, 2004). Se trata de una herramienta interesante porque vuelve a requerir la entrada subjetiva de los responsables de la toma de decisiones, y esta entrada subjetiva principal son las ponderaciones.

Hwang et al. (Yoon & Hwang, 1995), propone el TOPSIS basado en que la opción óptima esté lo más cerca posible y la opción ideal negativa lo más lejos posible. Y la metodología se representa con el siguiente diagrama de la figura 1 (Kolios, Mytilinou, Lozano-Minguez, & Salonitis, 2016)

La técnica TOPSIS puede ejecutarse mediante los siguientes pasos.

1. Construcción de la matriz de decisión: Partiendo de m alternativas $A_i, i=1, 2, \dots, m$ que serán evaluadas en función de los criterios $C_j, j=1, 2, \dots, n$.
2. Se obtienen los datos de rendimiento de las n alternativas en función de k criterios. Estas medidas deben ser estandarizadas por algún método de normalización.
 - a) Normalización distributiva.
 - b) Normalización ideal.

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^m (x_{ij})^2}}, j = 1, \dots, n, i = 1, \dots, m$$

Ecuación 15

3. Se desarrolla un conjunto de ponderaciones significativas para cada uno de los criterios. La base para estas ponderaciones puede ser alguna característica, pero puede ser alguna importancia relativa.

$$v = w x n_j = n_i = m$$

Ecuación 16

4. Las puntuaciones de los pesos se comparan con cada acción, ya sea una acción ideal o una acción virtual anti-ideal. Hay tres formas diferentes de definir estas acciones virtuales.

CAPÍTULO II. MARCO TEÓRICO

- a) Obtención del mejor y peor resultado en cada criterio de la matriz de decisión normalizada.

$$A^+ = \{v_1^+, \dots, v_n^+\} = \left\{ (\max_i v_{ij}, j \in J) (\min_i v_{ij}, j \in J) \right\} \text{ Ecuación 17}$$

$$A^- = \{v_1^-, \dots, v_n^-\} = \left\{ (\min_i v_{ij}, j \in J) (\max_i v_{ij}, j \in J) \right\} \text{ Ecuación 18}$$

- b) Asumir un punto ideal y anti-ideal absoluto, que se definen sin considerar las acciones del problema de decisión.
- c) Los puntos ideal y anti-ideal son definidos por el decisor.
5. Se calcula una medida de distancia para cada criterio, ideal y anti-ideal alternativo.

$$d_i^+ = \left\{ \sum_{j=1}^n (v_{ij} - v_j^+)^2 \right\}^{\frac{1}{2}} \text{ Ecuación 19}$$

$$d_i^- = \left\{ \sum_{j=1}^n (v_{ij} - v_j^-)^2 \right\}^{\frac{1}{2}} \text{ Ecuación 20}$$

6. Para cada alternativa se determina un coeficiente de proximidad igual a la distancia al anti-ideal dividida por la suma de la distancia al anti-ideal y la distancia al ideal.

$$R_i = \frac{d_i^-}{d_i^+ + d_i^-}, i = 1, \dots, m \text{ Ecuación 21}$$

7. Las alternativas obtenidas se ordenan maximizando el coeficiente de proximidad del paso 5

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN

Capítulo III. Metodología de la investigación

3.1. Introducción.

Esta investigación toma como base principal la información de 89 centros de salud y 5 hospitales generales pertenecientes al Instituto de Salud del Estado de Aguascalientes, en su Sistema de Consulta General y Consulta Externa Especializada el cual genera una base de datos de los pacientes consultados que es alimentada en la consulta diaria a través los diagnósticos del Catálogo Internacional de Enfermedades (CIE) versión 10.

Inicialmente se tomó la información del total de pacientes consultados seleccionando la información referentes a tres grupos de pacientes, los que fueron diagnosticados como probables con Diabetes Mellitus, diagnosticados como confirmados con Diabetes Mellitus, diagnosticados como No diabéticos, cuyas consultas fueron realizadas en los años 2015-2022, tomando información que no sea sensible a alguna circunstancia legal, así mismo esta información se preproceso para eliminar registros de pacientes con datos incompletos así como una reducción de las variables iniciales dejando sólo las variables consideradas de importancia para el estudio.

En esta investigación se tomaron únicamente datos no sensibles de los pacientes y las consultas, de personas que han sido consultados y confirmados en la Dirección de Servicios de Salud del Instituto de Servicios de Salud del Estado de Aguascalientes en su primer y segundo nivel de atención.

3.2. Objeto de Estudio.

La información de la base de datos fue proporcionada por el Instituto de Servicios de Salud del Estado de Aguascalientes, en coordinación con la Sub-Dirección de Informática y Estadística, se supervisó que la información para esta investigación fuera información no sensible de los pacientes y las consultas realizadas del año 2015 al 2022.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN

La clasificación de los diagnósticos asignados en consulta a los pacientes se realiza mediante el CIE10 mediante una codificación alfanumérica de 4 dígitos, de los cuales el primero es una letra y los siguientes tres son un número consecutivo. Se muestra el listado total de diagnósticos referentes a Diabetes en el Anexo A.

El insumo de datos de la base inicial es de aproximadamente 200,000 pacientes, de estos fueron seleccionados pacientes con probable diabetes quedando más de 10,000 pacientes. Después se rastrearon pacientes diagnosticados confirmados con diabetes mellitus. La población inicial se determinó con un cruce de datos alcanzando un total de 600 pacientes, Los cuales se filtraron y analizaron para esta investigación.

3.3. Criterios de Selección

3.3.1. Criterios de inclusión.

En esta investigación se consideran sólo datos de pacientes:

- Diagnosticados en consulta con Diabetes Mellitus mediante un Diagnóstico del CIE 10.
- Diagnosticados como probables diabéticos, diabéticos confirmados y no diabéticos consultados del año 2015 al 2022.
- Sin distinción de sexo, edad, estatura, medidas, etc.
- Con datos completos de la base de datos en las variables a analizar.
- Se incluyen consultas de pacientes considerados como sanos de acuerdo con el CIE 10.

3.3.2. Criterios de exclusión.

Quedan excluidos de esta investigación los pacientes:

- En consultas con diagnósticos diferentes a los de Diabetes Mellitus del CIE 10.
- Consultados fuera de los años 2015 al 2022.
- Con datos insuficientes o con datos nulos en la base de datos respecto de las variables a estudiar.
- Se descartan las consultas de pacientes con datos incorrectos y / o fuera de rango

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN

3.4. Definición de Variables.

Originalmente la base de datos de consulta conto con aproximadamente un total de 25 variables consideradas para la investigación (Tabla 8).

Tabla 8 Variables Iniciales de la Base de Datos

VARIABLES
Antec. Fam. Diabetes
Calle
Colonia
Complicación
Derechohabiencia
Edad
Estado
Estado Civil
Frecuencia Cardiaca
Frecuencia Respiratoria
Fuma
Glucemia
Índice de Masa Corporal
Indígena
Migrante
Municipio
Nivel trabajo social
Numero
Programa INSABI
Pulso
Religión
Sexo
Tensión Arterial Baja
Tensión Arterial Alta
Temperatura

Después de consultar las variables con expertos en diabetes e identificadas en la literatura, se determinaron como importantes las siguientes variables mostradas a continuación y de las cuales se determinaron nivel de importancia, posibles valores y desratización.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN

3.4. 1 Definición Constructiva.

- **Antecedentes Familiares de Diabetes:** es considerado un factor de riesgo muy importante en diabetes, este es determinado por el médico tratante a través de antecedentes familiares y determina si el antecedente es importante en hasta dos generaciones ascendentes.
- **Complicación:** es una variable basada en el historial del paciente determinada por la existencia de alguna complicación en los diagnósticos del paciente, por ejemplo, enfermedades cardíacas, cerebrales, de visión, renales, nerviosas, etc.
- **Derechohabiencia:** tipo de servicio o prestaciones sanitarias públicas el cual el paciente tiene derecho a recibir.
- **Edad:** edad del paciente en el momento de la consulta.
- **Estado:** información de ubicación como la colonia, el municipio y el estado de residencia del paciente.
- **Estado Civil:** condición de una persona según el registro civil en función de si tiene o no pareja y su situación legal respecto a esto.
- **Frecuencia Cardíaca:** es la frecuencia de vibraciones de las paredes de las arterias que se produce al pasar por estas una onda recurrente de sangre bombeada por la contracción ventricular.
- **Glucemia:** la cantidad de un azúcar o glucosa en una muestra de sangre.
- **Índice de Masa Corporal:** el índice de masa corporal (IMC) es una guía sencilla de la relación entre el peso y la talla que utilizada comúnmente en la identificación del sobrepeso y la obesidad en los adultos. Esta se calcula dividiendo el peso del individuo por el cuadrado de su tamaño en metros (kg/m^2).
- **Sexo:** es el género biológico del paciente (Masculino o Femenino).
- **Nivel de trabajo social o Deciles:** es un nivel de clasificación que asigna el personal de trabajo social este determina el nivel de aportación gubernamental que recibe el paciente.
- **Tensión Arterial Alta y Baja:** es la fuerza creada por el corazón, mantenida por la elasticidad arterial y regulada por las resistencias periféricas.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN

- **Temperatura:** es un signo vital que interpreta la armonía entre el calor que produce el metabolismo, acción muscular, alimentos y oxígeno como origen de energía que se pierde por medio de la piel, pulmones y evacuaciones del cuerpo. 3.4.2. Definición Operacional.
- **Antecedentes Familiares Diabetes:** es una variable que indica presencia o ausencia de familiares con diabetes 0,1.
- **Complicación:** es una variable que indica presencia o ausencia de una complicación de salud en el paciente 0,1.
- **Derechohabiciencia:** este dato es determinado por el departamento de trabajo social de la unidad médica u Hospital un número del 1 al 6.
- **Edad:** se calcula en base a los años cumplidos en el momento de la consulta.
- **Estado:** se determina por medio del catálogo de Colonias, Municipios y Estados establecido por el INEGI el cual es capturado por el departamento de archivo es un número del 1 al 32.
- **Estado Civil:** situación de un individuo ante el registro civil bajo la determinación de sí o no tiene pareja y su condición constitucional en relación con esto. Es un número del 1 al 7
- **Frecuencia Cardíaca:** es medida en el momento previo a la consulta y es el número de latidos por minuto.
- **Glucemia:** se miden por medio de una muestra de sangre ya sea en laboratorio o por dispositivo electrónico y tiras reactivas, se mide en mg/dl.
- **Índice de Masa Corporal:** $IMC = \text{Peso} / (\text{Talla})^2$
- **Sexo:** Masculino o Femenino.
- **Nivel de trabajo social o Deciles:** es un nivel de clasificación otorgado por el departamento de trabajo social el cual define el nivel de apoyo económico que recibe el paciente es un número del 1 al 10.
- **Tensión Arterial Alta y Baja:** se mide en mmHg y es el empuje aplicado por 1 milímetro (mm) de mercurio (Hg).
- **Temperatura:** es medida por medio de termómetro en un momento previo a la consulta.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN

3.5 Cálculo del Tamaño de la Muestra.

Para establecer la validez de nuestro Algoritmo de diagnóstico metaheurístico de aprendizaje automático, se llevó a cabo un cálculo de muestra que resultara representativo de la población de estudio: pacientes con la patología de Diabetes Mellitus en el estado de Aguascalientes.

A continuación, se muestra cómo se calculó el tamaño de muestra con la que se implementó del clasificador de redes neuronales artificiales.

$$n = \frac{N\sigma^2 Z^2}{e^2(N-1) + \sigma^2 Z^2} \quad \text{Ecuación 22}$$

Donde

$N=20,000$

$Z=1.96$ (suponiendo una confianza de 95%)

$\sigma= 0.5$

$e=0.05$ (suponiendo un error de 5%)

$$n = \frac{20000(0.5)^2(1.96)^2}{(0.05)^2(19999) + (0.5)^2(1.96)^2} \quad \text{Ecuación 23}$$

Por lo tanto, n puede ser de un tamaño de 400 Pacientes.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

Capítulo IV. Marco de trabajo para creación y puesta en marcha de sistema metaheurístico de aprendizaje automático para factores de diabetes mellitus

El marco de trabajo propuesto para el desarrollo de la presente investigación se compone de los siguientes elementos fundamentales:

1. ACERCAMIENTO CON LA FUENTE DE DATOS: se dialogó y estableció un convenio de confidencialidad y manejo de información con la subdirección de informática y estadística del Instituto de servicios de Salud del Estado de Aguascalientes (89 centros de salud y 5 hospitales generales los cuales cuentan con Expediente Clínico Electrónico ECE.)
2. COLECCIÓN Y PREPARACIÓN DE INFORMACIÓN: se obtuvo la información de respaldos actuales de las 94 unidades para ser montados en 2 servidores de MS SQL Server 2008 R2.

Se diseñaron y crearon de consultas específicas para la exploración de cada una de las bases de datos mediante la selección de pacientes con la patología de Diabetes Mellitus Según diagnóstico y clave del CIE10.

Se extrajo y almacenó la información en una sola base de datos montada en MS SQL Server 2008 R2 para mejor accesibilidad y tratamiento.

Se seleccionó la información de los grupos de pacientes con el diagnóstico de probable Diabetes Mellitus, confirmados Diabetes Mellitus y sin diagnóstico de Diabetes Mellitus apoyados en el CIE 10 por capítulo, grupo, categoría y subcategoría mediante el asesoramiento de expertos en la materia (ORGANIZACIÓN PANAMERICANA DE LA SALUD, 2015).

Basados en la literatura y los catálogos del INEGI se discretizó la información de la base de datos de 25 variables disponibles.

Se implementó la selección de subconjuntos de características para reducir el número de factores del problema.

3. REVISIÓN DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO: se estudió, desarrollo y se implementaron diversas heurísticas y metaheurísticas, previo al diseño de Algoritmo Metaheurístico Híbrido. Se estudiaron las técnicas de Ingenuo de Bayes y de Redes Neuronales artificiales profundas.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

4. DISEÑO E IMPLEMENTACIÓN DE ALGORITMO DE DIAGNOSTICO DE METAHEURÍSTICO DE APRENDIZAJE AUTOMÁTICO PARA FACTORES DE DIABETES MELLITUS: se diseñó y experimento el Algoritmo Metaheurístico Híbrido para clasificación e identificación de factores asociados a diabetes.
Se compararon los resultados de precisión tanto de las redes neuronales como del ingenio de bayes desarrollados.
Se seleccionó el mejor resultado de precisión de los clasificadores mencionados.
Se experimentó con la propuesta de Algoritmo Metaheurístico.
5. DISEÑO E IMPLEMENTACIÓN DE SISTEMA DE DIAGNÓSTICO DE METAHEURÍSTICO DE APRENDIZAJE AUTOMÁTICO PARA FACTORES DE DIABETES MELLITUS: se diseñó e implemento el sistema de diagnóstico metaheurístico de aprendizaje automático.

Se programó dentro del expediente clínico electrónico (ECE) el código para invocar el proceso de clasificación.
Se programaron y almacenaron los programas y bases de datos correspondientes para el manejo de las variables y los resultado del clasificador.
Se programo la alerta de Diabetes Mellitus en el tablero de alertas del ECE.
6. PROPUESTA DE UBICACIÓN DE CENTROS DE ATENCIÓN: se planteó la creación de un algoritmo K-Means modificado para determinar los puntos de concentración de pacientes diagnosticados con diabetes mellitus, agrupar por su cercanía a las jurisdicciones correspondientes e incluso se propuso nuevas jurisdicciones sanitarias para tener una mejor distribución de la población y reducir las distancias de centros de atención hacia su jurisdicción responsable.
7. Se desarrolló y experimento el procesamiento de software de clusterización y la selección inteligente multicriterio TOPSIS.
Se desarrolló y experimento con el algoritmo de K-Means modificado para centroides fijos y móviles, obteniendo un nuevo centroide. Se implementó la técnica de TOPSIS para selección inteligente multicriterio.
8. VALIDACIÓN Y DISCUSIÓN DE RESULTADOS: etapa en la que se validan y discuten los resultados, respecto al objetivo general, objetivos específicos e hipótesis.

A continuación, se presenta un diagrama a bloques detallando los componentes del Marco de Trabajo:

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

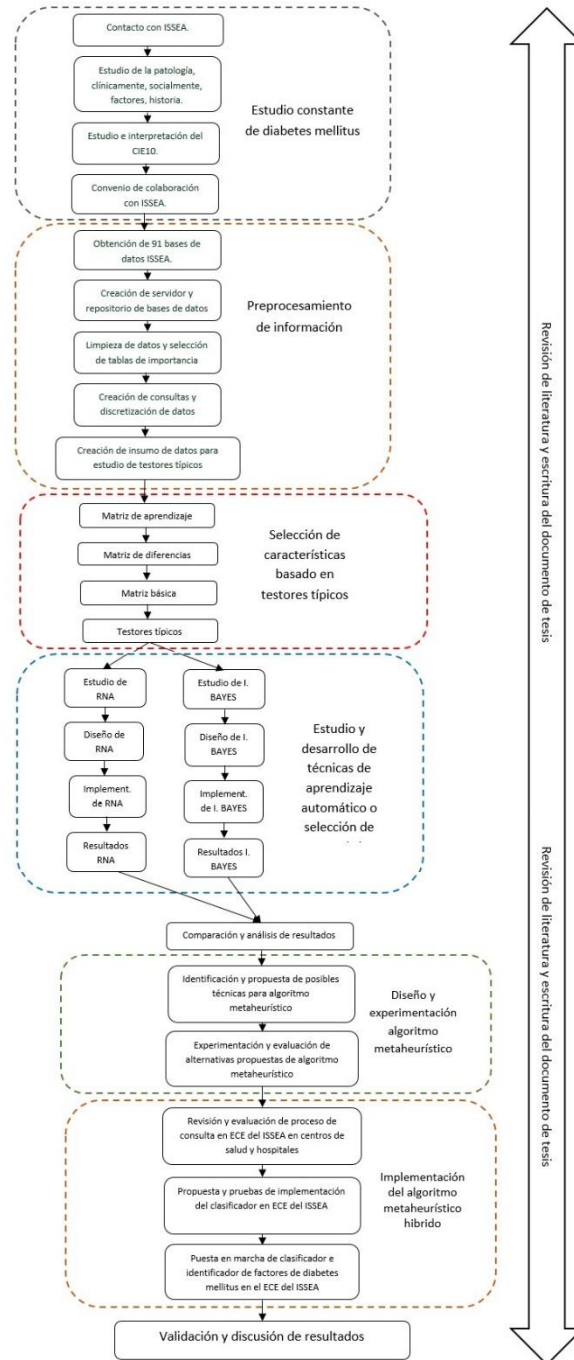


Figura 12 Marco de Trabajo de la investigación

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

4.1. Marco de trabajo de algoritmo metaheurístico de aprendizaje automático.

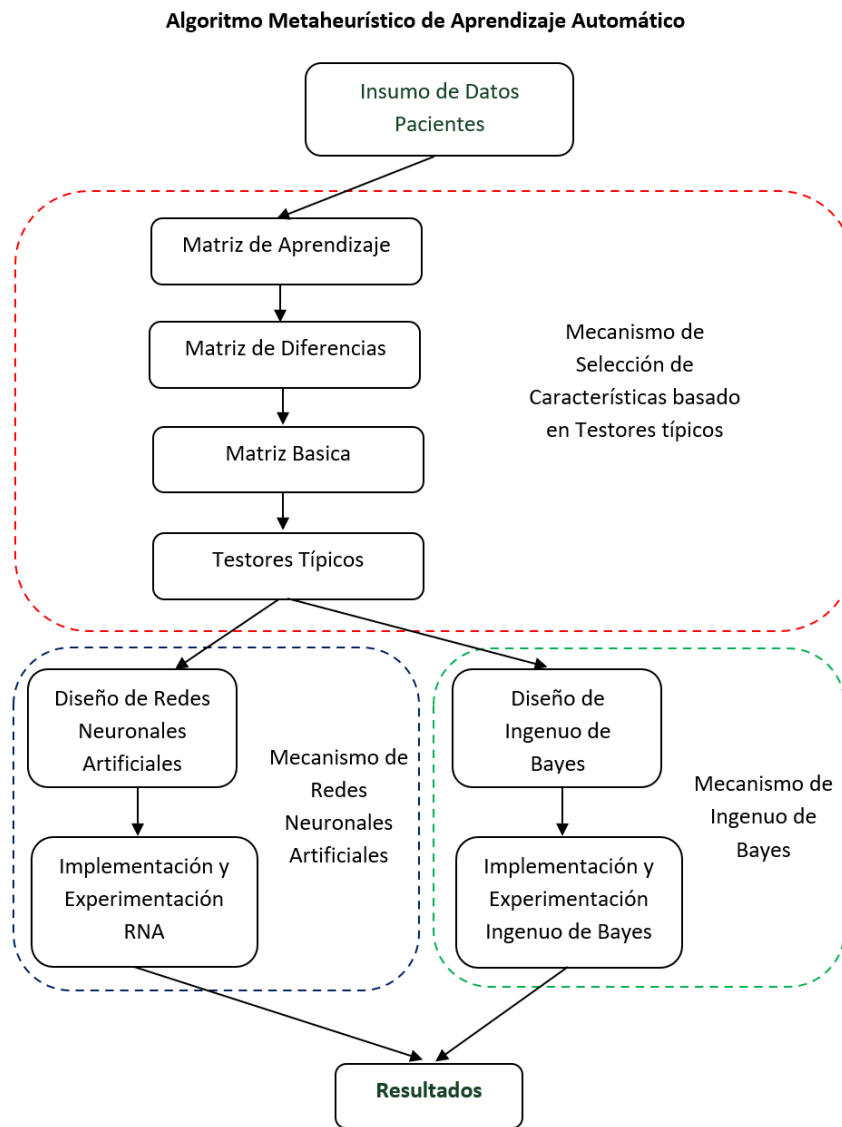


Figura 13 Marco de Trabajo de algoritmo metaheurístico de aprendizaje automático

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

4.2. Preparación de la información.

La etapa de preparación de la información consta de una limpieza y discretización de la base de datos.

Inicialmente, se eliminaron los elementos repetidos, con errores de captura o con datos nulos. Posteriormente, se ubicaron los límites de la información mediante el diseño de consultas para finalmente, discretizar con apoyo de la literatura de Diabetes Mellitus como se muestra a continuación.

Para la edad de las personas establece (Mansilla A, 2010) una clasificación de las etapas de la vida (Tabla 9):

Tabla 9 Clasificación de Etapas de la Vida

Clasificación de etapas de la vida.		
ETAPA PRENATAL	GESTACIÓN-NACIMIENTO	menos de 0 años
ETAPA FORMATIVA	NIÑEZ-ADOLESCENCIA	(0 a 17 años)
ETAPA LABORAL	JUVENTUD	18 a 24 años
	ADULTEZ	25 a 64 años
ETAPA JUBILAR	SENECTO	65 a 74 años
	ANCIANOS	75 a 95 o más años

Tomando en cuenta lo anterior y considerando que el hospital no cuenta con atención a pacientes menores de edad para esta variable se establecieron las siguientes reglas de discretización, **Edad:** 4 niveles, 1-joven, 2- Adulto, 3- Senecto, 4-Anciano (Tabla 10).

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

Tabla 10 Discretización de Edad

EDAD	
1	JOVEN <24 AÑOS
2	ADULTO >=24 AÑOS <65 AÑOS
3	SENECTO >=65 AÑOS <75 AÑOS
4	ANCIANO >=75 AÑOS

En la definición de la variable de glucemia se consideró como guía el nivel que determina el ISSEA dentro del sistema de consulta, el cual establece **Glucemia:** miligramos por decilitro (mg/dl) de glucosa en sangre, esta variable se determinaron 3 niveles, 1-Sin Riesgo, 2- Normal en control, 3-Con Riesgo (Tabla 11 se modificaran al terminar estructura).

Tabla 11 Discretización de Glucemia

GLUCEMIA	
1	SIN RIESGO <100
2	NORMAL EN CONTROL >=100<=140
3	CON RIESGO >140

Dentro de la clasificación de Hipertensión Arterial que (Parra Martínez, 2014) menciona, presenta la clasificación de niveles de la Hipertensión arterial (Tabla 12) la cual fue promulgada por La Sociedad Española de Hipertensión-Liga Española para la Lucha contra la Hipertensión Arterial.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

Tabla 12 Clasificación de Hipertensión Arterial

Clasificación de niveles de HTA			
		Sistólica	Diastólica
1	OPTIMA	<120	<80
2	NORMAL	120-129	80-84 (prehipertensión)
3	NORMAL-ALTA	130-139	85-89 (prehipertensión)
4	GRADO 1	140-159	90-99 (ligera)
5	GRADO 2	160-179	100-109 (moderada)
6	GRADO 3	180-209	110-119 (severa)
7	GRADO4	210	120

Para discretizar los niveles de tensión arterial se procedió a separar las lecturas de tensión Alta y tensión Baja y establecer 7 niveles de **Tensión Arterial**: 1-Tensión óptima, 2- Tensión Normal, 3- Tensión Normal Alta, 4-Hipertensión Grado 1, 5-Hipertensión Grado 2, 6-Hipertensión Grado 3, 7- Hipertensión Grado 4 (Tabla 13).

Tabla 13 Discretización de Tensión Arterial

TENSIÓN ALTA Y BAJA			
		ALTA	BAJA
1	TENSIÓN OPTIMA	<120	<80
2	TENSIÓN NORMAL	120-129	80-84
3	TENSIÓN NORMAL-ALTA	130-139	85-89
4	HIPERTENSIÓN GRADO 1	140-159	90-99
5	HIPERTENSIÓN GRADO 2	160-179	100-109
6	HIPERTENSIÓN GRADO 3	180-209	110-119
7	HIPERTENSIÓN GRADO 4	210	120

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

Respecto a la variable de frecuencia cardiaca se encontró la Clasificación de frecuencia cardiaca en reposo presentada por (Gennaro, 1983) Tabla 14, que establece los siguientes niveles:

Tabla 14 Frecuencia Cardiaca

Clasificación	Frecuencia Cardiaca en Reposo (latidos / Minuto)
Excelente	≤ 55
Muy Buena	56 - 65
Buena	66 - 70
Promedio	71 - 80
Pobre	81 - 95
Muy Pobre	≥ 96

Estableciendo 6 niveles para discretizar la variable de **Frecuencia Cardiaca**: 1- Excelente, 2- Muy Buena, 3-Buena, 4-Promedio, 5-Pobre, 6- Muy Pobre en tabla 15.

Tabla 15 Discretización de Frecuencia

Frecuencia Cardiaca		
1	EXCELENTE	≤ 55
2	MUY BUENA	56 - 65
3	BUENA	66 - 70
4	PROMEDIO	71 - 80
5	POBRE	81 - 95
6	MUY POBRE	≥ 96

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

Basados en literatura tal como lo presenta (Mondelo, Enrique Torada, Comas Uriz, Castejón Vilella, & Bartolomé Lacambra, 1999) menciona que la temperatura corporal es producto de la armonía entre las ganancias y pérdidas del calor ubicado en el interior de un microsistema, y establece la siguiente escala de temperatura corporal en tabla 16.

Tabla 16 : Escala de Temperatura Corporal

Escala de la temperatura corporal	
44°C	Golpe de calor
42°C	*convulsiones, coma
41°C	*piel caliente y seca
40°C	Hiperpirexia
38°C	Intervalo aproximado de temperatura normal
36°C	
34°C	Sensación de frío
33°C	Hipotermia
32°C	*bradicardia, hipotensión
30°C	*somnolencia, apatía
28°C	*muscultura rígida
26°C	Límite inferior de supervivencia

Basado en esta escala se determinó que la Temperatura: Medida en grados centígrados se puede dividir en 6 niveles, 1- Límite de Supervivencia, 2- Hipotermia, 3-Frío, 4-Normal, 5- Hiperpirexia, 6 Golpe de Calor. La escala de discretización quedó de la siguiente manera en tabla 17.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

Tabla 17 Discretización de Temperatura Corporal

TEMPERATURA		
1	LIMITE DE SUPERVIVENCIA	<28°C
2	HIPOTERMIA	>=28°C <33°C
3	FRIO	>=33°C <36°C
4	NORMAL	>=36°C <38°C
5	HIPERPIREXIA	>=38°C <44°C
6	GOLPE DE CALOR	>=44°C

Se tomaron de referencia de los catálogos del ISSEA los siguientes valores para las siguientes variables:

- **Antecedentes familiares Diabetes:** 0- No existencia de antecedentes, 1- Existencia de antecedentes.
- **Complicación:** 0- No existencia de complicación, 1- Existencia de complicación.
- **Nivel Trabajo Social o Deciles** es una conversión de los niveles determinados para cada paciente por el departamento de trabajo social del hospital, con los cuales se establecieron 3 niveles, 1.- El paciente tiene apoyo por parte de seguro popular, 2.- El paciente cuenta con apoyo de gobierno estatal y 3.- El paciente no cuenta con apoyo.
- **Estado Civil:** Estado civil del paciente 1-Soltero, 2-Casado, 3-Unión Libre, 4-Separado, 5-Divorciado, 6-Viudo, 7-Desconocido.
- **Sexo:** 1- Masculino, 2 – Femenino
- **Derechohabiciencia:** 0-Población Abierta, 1-Seguro Popular, 2-IMSS, 3-ISSSTE, 4-PEMEX, 5-SECMAR, 6-Otra y 7-SMNG.
- **Índice de Masa Corporal:** 1-Desnutrido, 2-Normal, 3-Sobre Peso, 4- Obesidad.
- **Estado** Respecto a esta variable se respetaron los catálogos nacionales establecidos por el INEGI y utilizados por el ISSEA.

A partir de esta clasificación y discretización se generó la siguiente Tabla de todas las

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO IV MARCO DE TRABAJO

SMAADM

variables con nivel, dominio de función y criterio de comparación en tabla 18.

Tabla 18 Dominio y Criterio de Comparación de Variables

VARIABLE	NIVEL	DOM. DE FUNC.	CRIT. DE COMP.
ANT. FAM. DIABETES	CLÍNICO	1-2	Igualdad Estricta
COMPLICACIÓN	CLÍNICO	1-2	Igualdad Estricta
DECILES	POBLACIONAL	1,2,3	Igualdad Estricta
DERECHOHABIENCIA	POBLACIONAL	1,2,3,4,5,6,7,8	Igualdad Estricta
ESTADO	POBLACIONAL	1-32(CATALOGO)	Igualdad Estricta
EDAD	CLÍNICO	1,2,3,4	Igualdad Estricta
ESTADO CIVIL	POBLACIONAL	1,2,3,4,5,6,7	Igualdad Estricta
FC	CLÍNICO	1,2,3,4,5,6	Igualdad Estricta
GLUCEMIA	CLÍNICO	1,2,3	Igualdad Estricta
IMC	CLÍNICO	1,2,3,4	Igualdad Estricta
SEXO	CLÍNICO	1,2	Igualdad Estricta
T.ALTA Y T.BAJA	CLÍNICO	1,2,3,4,5,6,7	Igualdad Estricta
TEMPERATURA	CLÍNICO	1,2,3,4,5,6	Igualdad Estricta

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO V IMPLEMENTACIÓN DEL SISTEMA SMAADM

Capítulo V. Implementación del Sistema SMAADM

En este capítulo se presenta el diseño e implementación del sistema metaheurístico de aprendizaje automático para factores de diabetes mellitus el cual está basado en redes neuronales artificiales y testores típicos y fue implementado dentro del sistema de consulta del ECE.

El desarrollo de las redes neuronales implementadas se llevó a cabo en lenguaje python utilizando las librerías de Tensorflow y Keras.

Este sistema clasificador determina una de las tres clases partiendo de la información almacenada en el sistema de consultas de ISSEA y propone un diagnóstico temprano al momento de la consulta en un centro de salud.

Es sistema pronostica las clases de:

- Diabético
- Prediabético
- No diabético.

A continuación, se presenta el diagrama del funcionamiento del sistema metaheurístico de aprendizaje automático para factores de diabetes mellitus.

CAPÍTULO V IMPLEMENTACIÓN DEL SISTEMA SMAADM

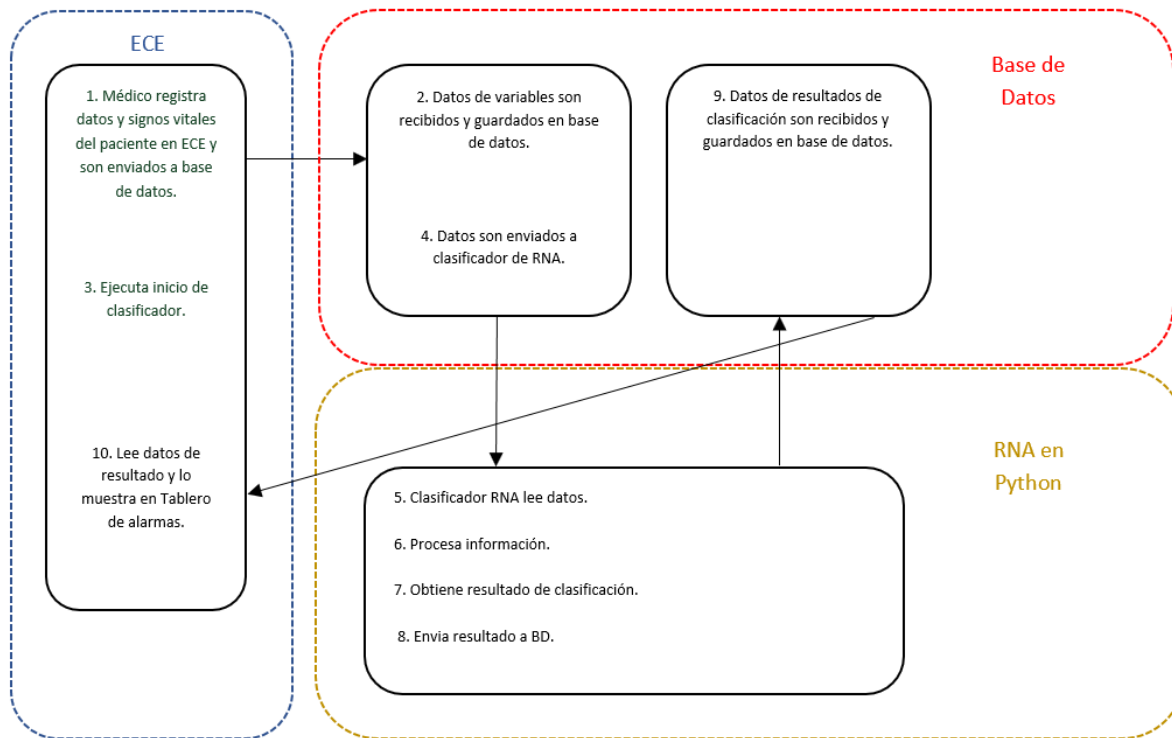


Figura 14 implementación del sistema SMAADM

Inicialmente, se realizó el guardado del modelo ya entrenado de la red neuronal en el archivo Prediccion.h5. Después, se implementó y programó un proceso en el cual se realiza un envío de los datos del paciente a una base de datos para predicciones que fue creada en el servidor de datos de la unidad de salud, esto ocurre cuando el médico termina el llenado de datos de nota médica en el sistema de Consulta del Expediente Clínico Electrónico (ECE) del ISSEA el cual está desarrollado en Delphi 7, en este mismo proceso se activa la ejecución de programa almacenado en servidor de datos el cual ejecuta un programa que realiza la predicción de todos y a cada una de las predicciones pendientes en la base de datos de predicciones, de esta manera tenemos el resultado de la predicción del paciente consultado en turno. Este resultado se muestra en el módulo de alertas en el sistema de consulta para ser considerado en el tablero de selección de diagnósticos CIE10. A continuación, en la figura 16, se muestra cómo se implementa la alerta o sugerencia de considerar paciente diabético dentro del sistema de ECE del ISSEA.

CAPÍTULO V IMPLEMENTACIÓN DEL SISTEMA SMAADM

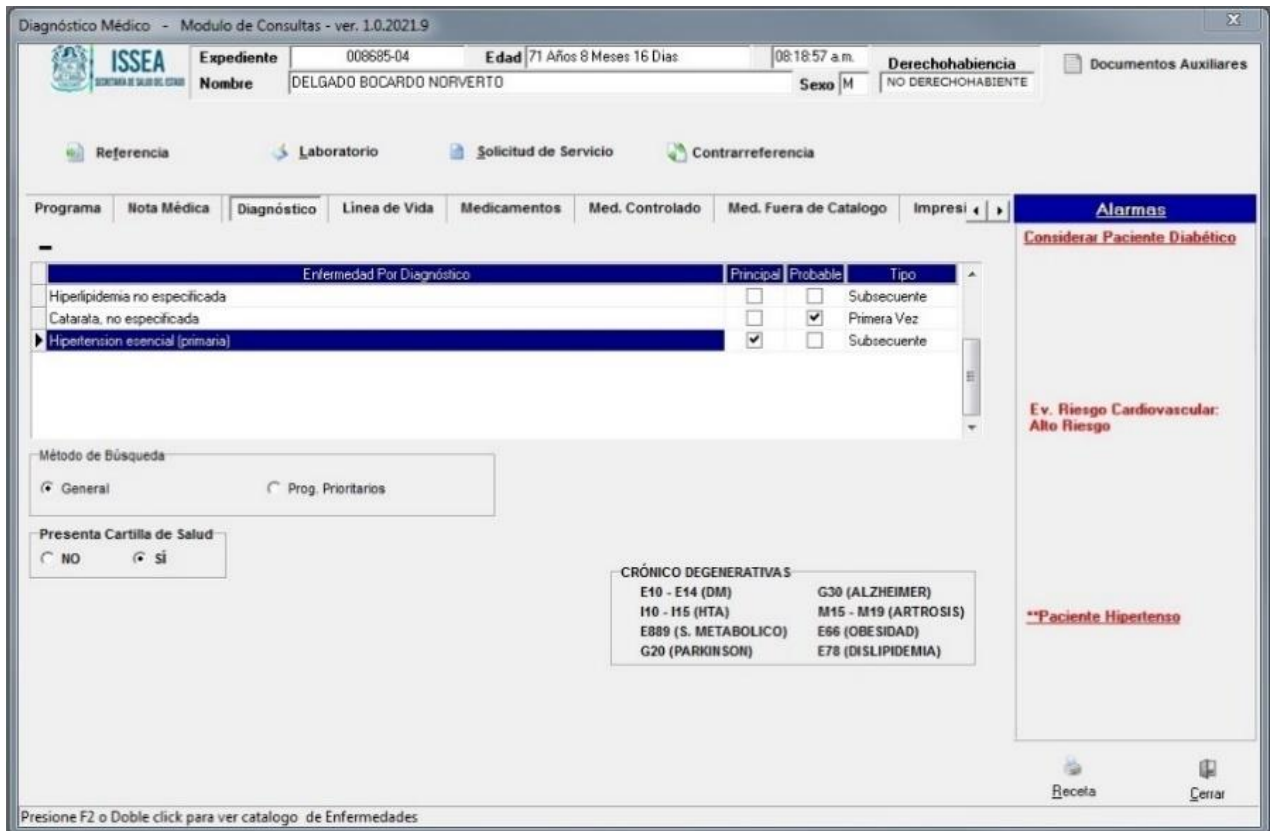


Figura 15 Alerta de considerar paciente diabético

A continuación, en la figura 17, veremos cómo se implementó la alerta o sugerencia de considerar paciente prediabético dentro del sistema de ECE del ISSEA.

CAPÍTULO V IMPLEMENTACIÓN DEL SISTEMA SMAADM

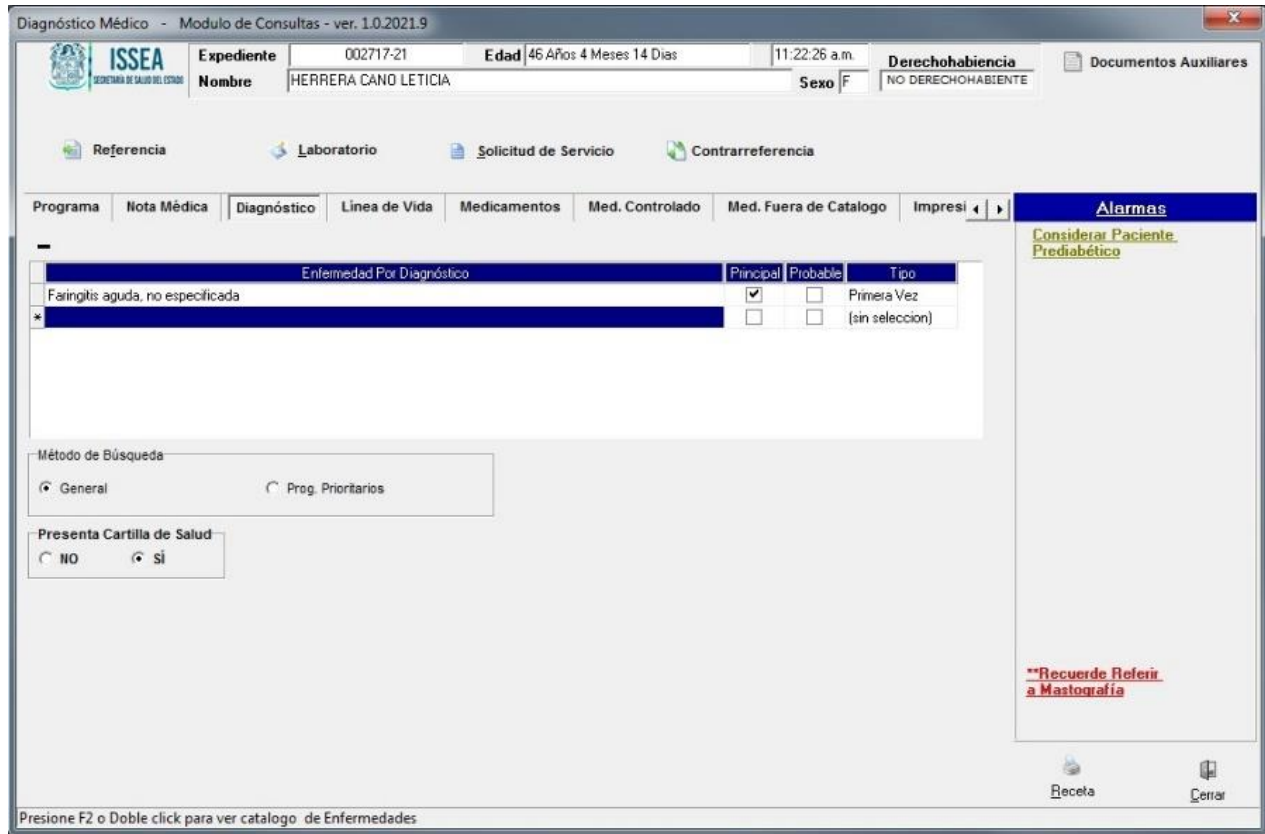


Figura 16 Alerta o sugerencia de considerar paciente prediabético

A continuación, en la figura 18, veremos cómo se implementa la alerta o sugerencia de considerar paciente no diabético dentro del sistema de ECE del ISSEA.

CAPÍTULO V IMPLEMENTACIÓN DEL SISTEMA SMAADM

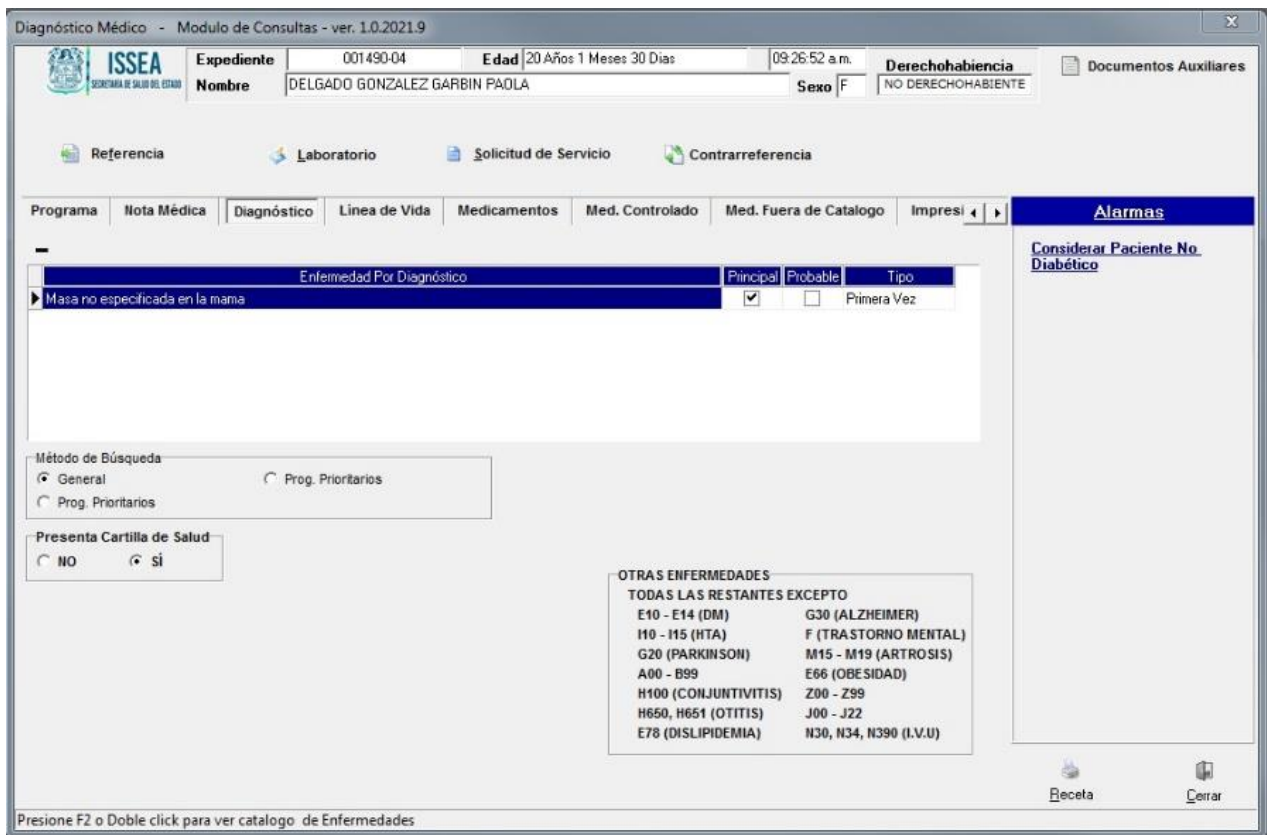


Figura 17 Alerta o sugerencia de considerar paciente no diabético

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Capítulo VI. Propuesta de reubicación de jurisdicciones

6.1 Diseño e implementación de algoritmo K-Means modificado

En este capítulo, se presenta una propuesta que busca reducir las distancias lineales y por carretera del centro de salud con respecto a su jurisdicción sanitaria de responsabilidad. Actualmente la distancia máxima de un centro de salud con respecto a su jurisdicción sanitaria es de aproximadamente 70 km por carretera y 54 kilómetros lineales.

A continuación, en la figura 19 se muestra un diagrama del procedimiento de reubicación de jurisdicciones y reagrupación de unidades de salud.

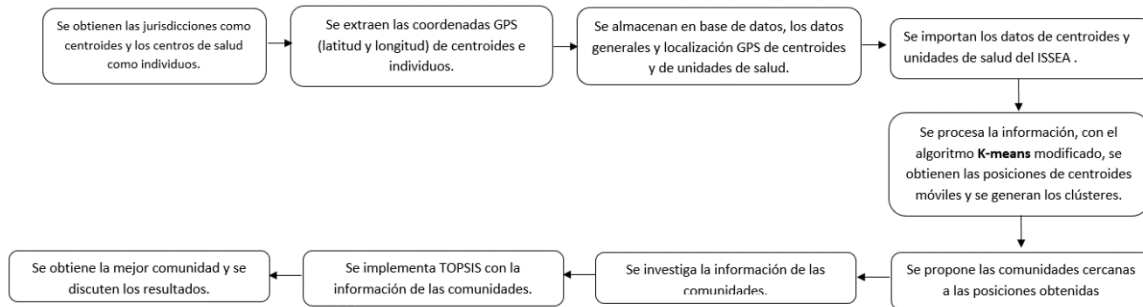


Figura 18 Procedimiento de reubicación de jurisdicciones y reagrupación de unidades de salud

Se desarrolló una versión de K-means con 3 centroides fijos y para cada una de las propuestas presentadas se trabajó con 1, 2 y 3 centroides variables. Los centroides fijos utilizan las coordenadas de longitud y latitud de las tres jurisdicciones sanitarias que tiene el ISSEA y los centroides variables representan las nuevas jurisdicciones sanitarias 4, 5 y 6, el cual se muestra a continuación.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Algoritmo K-means modificado.

m: cantidad total de centroides.

n: cantidad de centroides móviles.

1. IMPORTA individuos.
2. IMPORTA centroides fijos y móviles.
3. MIENTRAS la matriz de grupos anterior diferente con la matriz de grupos nueva.
 - a) LIMPIA matriz de grupos anterior con ceros.
 - b) LIMPIA matriz de centroides nuevos.
 - c) RECORRE todas las líneas de individuos y COMPARA con centroides y GENERA matriz de distancias resultantes.
 - d) RECORRE todas las líneas de distancias y COMPARA cual es la menor y GENERA matriz de grupos nueva.
 - e) ESTABLECE centroides fijos $1-m$.
 - f) RECORRE todas las líneas de grupos $(m+1)-n$ y SUMA los que son 1 en sus respectivos individuos y GENERA nueva posición de centroides móviles.
 - g) COMPARA matriz de grupos anterior con matriz de grupos nueva.
4. PREGUNTA, ¿son diferentes la matriz de grupos anterior con la matriz de grupos nueva?
5. SI.
 - a) LIMPIA matriz de grupos nueva con ceros.
 - b) IGUALA matriz de centroides anteriores a matriz de centroides nuevos.
 - c) LIMPIA matriz de centroides nuevos.
 - d) REPITE 3 y 4.
6. NO.
 - a) TERMINA.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Como resultado del algoritmo, se propone la inclusión de la jurisdicción 4 como centroide nuevo, contando así con 4 agrupaciones en total, esta decisión es de carácter económico (pocos centroides o jurisdicciones). Para la ubicación de la nueva jurisdicción se proponen 4 comunidades o ubicaciones posibles, ya que son comunidades cercanas al punto determinado por el algoritmo y están habilitadas con servicios básicos. Las ubicaciones son: La Luz, Terremoto, Pilotos y Amarillas de Esparza. La distancia del punto “Jurisdicción 4” determinado por el algoritmo (22.037063, -102.003551) al que llamaremos punto de atención o PDA a comunidades cercanas con los servicios necesarios para operar una oficina Jurisdiccional son:

1. PDA a La luz 4.34km con 870habitantes (figura 20).
2. PDA a El Terremoto 8.46km con 354 habitantes (figura 21).
3. PDA a Pilotos 4.45km 1282 con habitantes (figura 22).
4. PDA a Amarillas de Esparza1.31km con 808 habitantes (figura 23).

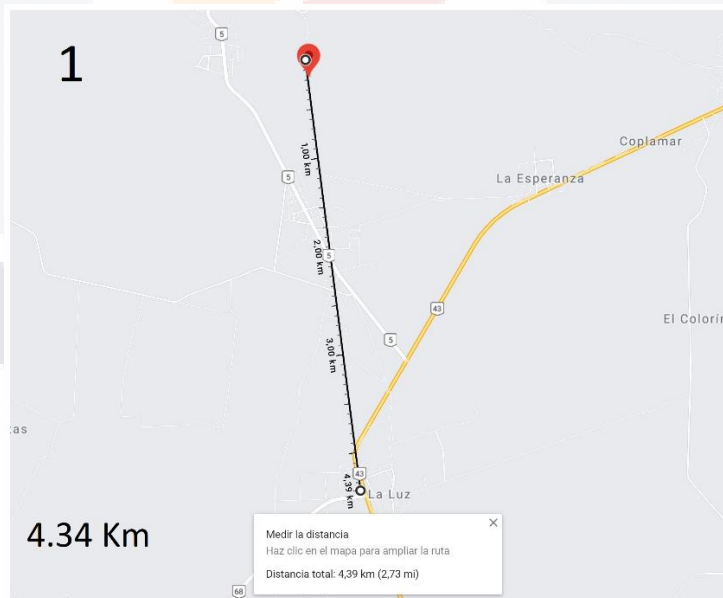


Figura 19 Localización de posible ubicación PDA 1

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

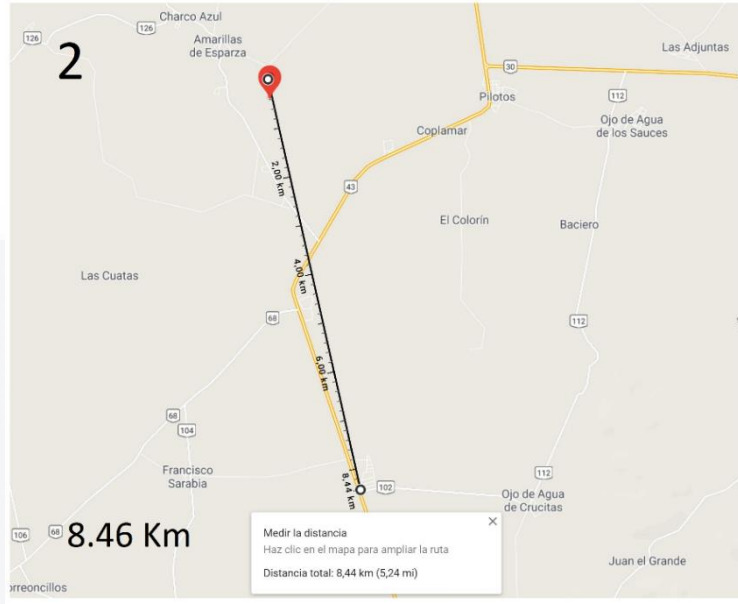


Figura 20 Localización de posible ubicación PDA 2

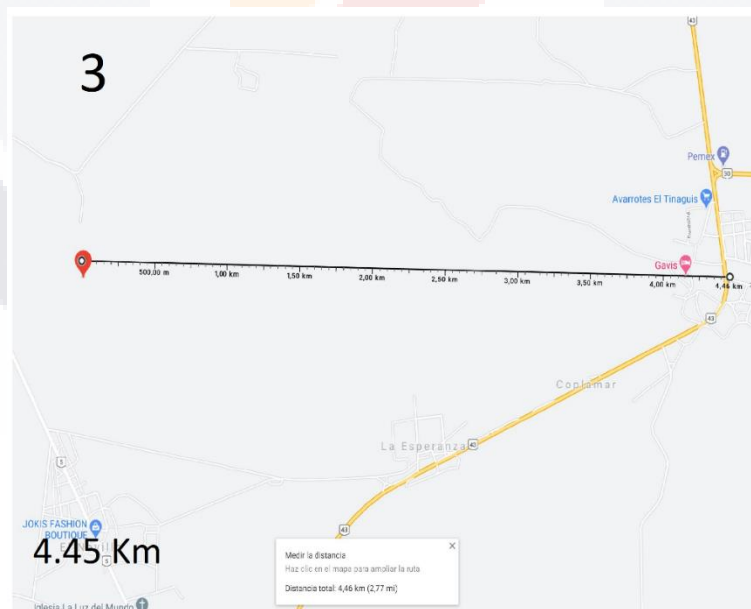


Figura 21 Localización de posible ubicación PDA 3

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

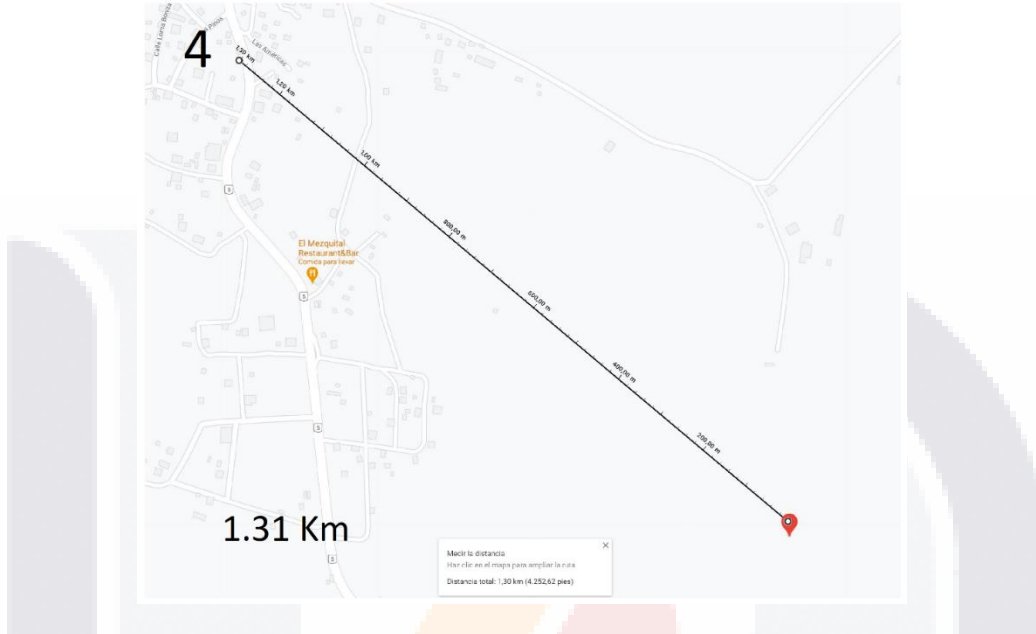


Figura 22 Localización de posible ubicación PDA 4

Actualmente, los 89 centros de salud atienden a un total de 658,904 pacientes registrados distribuidos de la manera mostrada en la tabla 19.

Tabla 19 Distribución Actual de Población

Jurisdicción	Unidades	Probables diabetes	Confirmados diabetes	Total, Pacientes
I	44	25,430	26,933	461,058
II	30	8,172	8,177	126,633
III	14	6,812	6,103	71,213
Todas	89	40,414	41,213	658,904

En la figura 24, se presenta la distribución actual de centros de salud por cada jurisdicción, de color azul los pertenecientes a la jurisdicción 1, en color naranja los pertenecientes a la jurisdicción 2 y en color verde los pertenecientes a la jurisdicción 3.

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

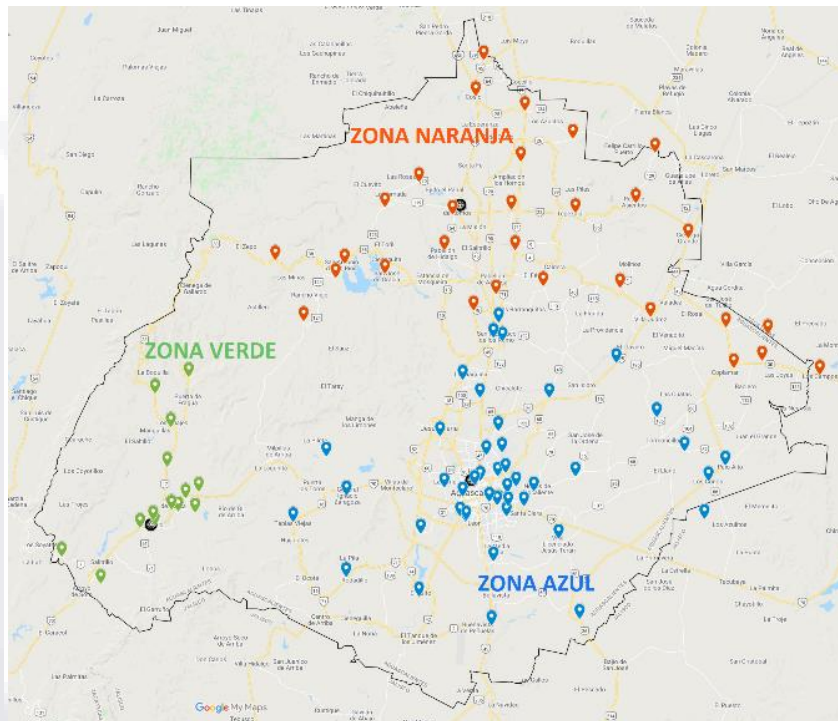


Figura 23 Distribución y ubicación actuales de centros de salud ISSEA

La propuesta de atención a pacientes registrados para cada una de las jurisdicciones quedaría redistribuida como se presenta en la tabla 20.

Tabla 20 Distribución de Población Propuesta

Jurisdicción	Unidades	Probables diabetes	Confirmados diabetes	Total, Pacientes
I	35	23,868	25,105	430,463
II	26	6,609	6,530	91,660
III	14	6,812	6,103	71,213
IV	14	3,125	3,475	65,568
Todas	89	40,414	41,213	658,904

En la figura 25, se presenta la propuesta de distribución de los 89 centros de salud en 4 jurisdicciones, en color azul los pertenecientes a la jurisdicción 1, en color naranja los pertenecientes

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

a la jurisdicción 2, en color verde los pertenecientes a la jurisdicción 3 y en color violeta los pertenecientes a la nueva jurisdicción 4.

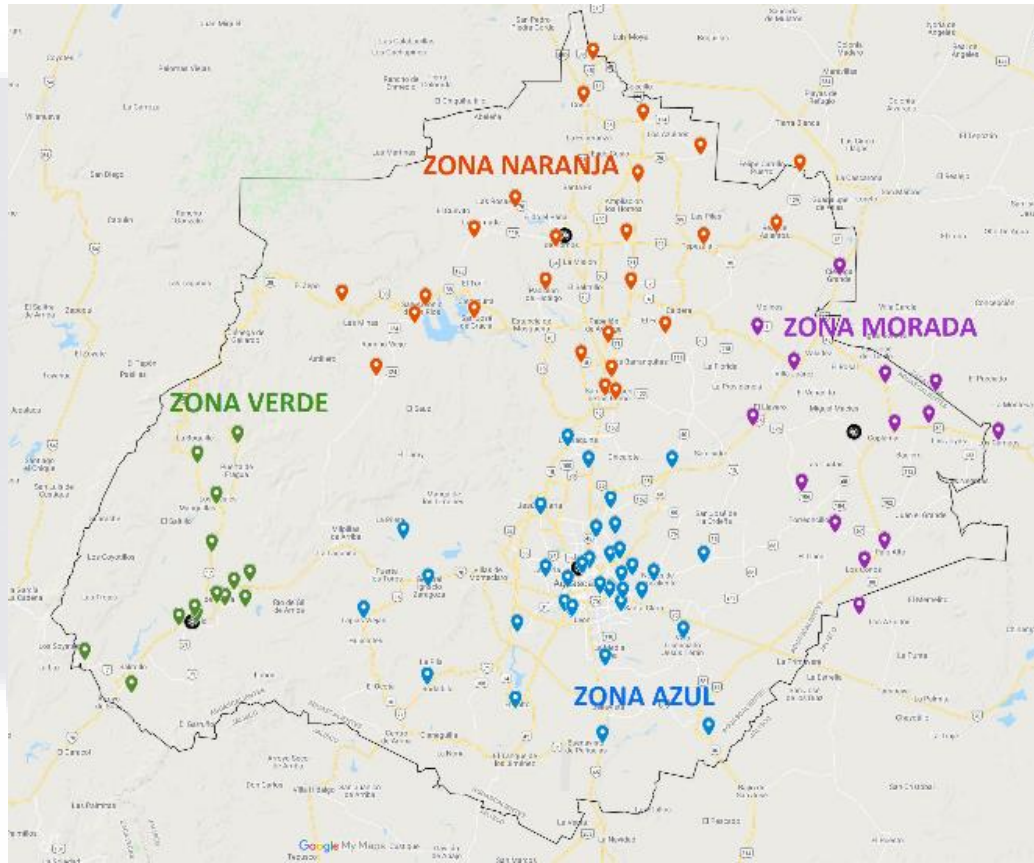


Figura 24 Nueva agrupación propuesta de centros de salud ISSEA

Como resultado del algoritmo se propuso la inclusión de la jurisdicción sanitaria 4 como nuevo centroide, teniendo así 4 clúster en total, La distancia del punto de la nueva jurisdicción sanitaria 4 que determinó el algoritmo fue la posición **(22.037063, -102.003551)** que nombraremos punto de atención o POA a las comunidades cercanas con los servicios necesarios para operar una oficina Jurisdiccional.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

6.2 Diseño e implementación de TOPSIS

Esta investigación cuenta con el apoyo de un maestro en planeación urbana que seleccionó 6 localidades cercanas (10 km de distancia máxima) y las variables importantes a utilizar en el TOPSIS, las localidades se enlistan en la tabla 21.

Tabla 21 Posibles localidades

Número	Localidad	Municipio
1	El Terremoto	El Llano
2	La Luz	El Llano
3	Santa Rosa	El Llano
4	Lic. Jesús Terán	El Llano
5	Pilotos	Asientos
6	Amarillas de Esparza	Asientos

En este experimento se aplicó la técnica de análisis TOPSIS, mediante la cual se realizó una clasificación de acuerdo con los criterios que se propusieron al inicio junto con el peso de cada uno de ellos que fueron seleccionados por el experto en urbanismo y sobre estas consideraciones se realizaron una serie de pasos para lograr una calificación que nos diera la solución ideal al problema. En las siguientes tablas 22 y 23 se presentan los datos obtenidos de los catálogos determinados por el INEGI, así como de los datos del SCITEL calificados por el experto en urbanismo, esta tabla incluye los pesos determinados para cada característica.

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Tabla 22 Matriz de criterios de decisión 1

Número	Localidad	1. Total habitantes	2. Número de viviendas	3. Servicios Electricidad	4. Servicios de Drenaje	5. Servicios de Agua Potable	6. Población con servicios de salud
1	El Terremoto	6	7	6	5	7	6
2	La Luz	5	5	7	5	5	7
3	Santa Rosa	7	5	9	7	5	9
4	Licenciado Jesús Terán	6	7	5	6	7	5
5	Pilotos	8	7	9	8	7	9
6	Amarillas de Esparza	7	6	5	7	6	5
.	Pesos	0.10	0.20	0.05	0.05	0.05	0.05

Tabla 23 Matriz de criterios de decisión 2

Número	Localidad	7. Tamaño de localidad	8. Distancia por carretera	9. Distancia lineal	10. Carreteras y caminos	11. Existencia de infraestructura ISSEA
1	El Terremoto	5	7	6	5	7
2	La Luz	5	5	7	5	5
3	Santa Rosa	7	5	9	7	5
4	Licenciado Jesús Terán	6	7	5	6	7
5	Pilotos	8	7	9	8	7
6	Amarillas de Esparza	7	6	5	7	6
.	Pesos	0.10	0.10	0.10	0.10	0.10

Se obtuvieron los valores normalizados y una medida del valor correcto, luego se procedió a evaluar estos resultados con el peso dado por el “experto en planificación y urbanización” y se obtuvo una matriz con los mismos valores ya normalizados en todos los criterios, la cual se presenta en las Tablas 24 y 25 a continuación.

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Tabla 24 Matriz normalizada 1

Número	Localidad	1. Total habitantes	2. Número de viviendas	3. Servicios Electricidad	4. Servicios de Drenaje	5. Servicios de Agua Potable	6. Población con servicios de salud
1	El Terremoto	0.3728	0.4586	0.3482	0.3175	0.4586	0.3482
2	La Luz	0.3107	0.3276	0.4062	0.3175	0.3276	0.4062
3	Santa Rosa	0.4350	0.3276	0.5222	0.4445	0.3276	0.5222
4	Licenciado Jesús Terán	0.3728	0.4586	0.2901	0.3810	0.4586	0.2901
5	Pilotos	0.4971	0.4586	0.5222	0.5080	0.4586	0.5222
6	Amarillas de Esparza	0.4350	0.3931	0.2901	0.4445	0.3931	0.2901

Tabla 25 Matriz normalizada 2

Número	Localidad	7. Tamaño de localidad	8. Distancia por carretera	9. Distancia lineal	10. Carreteras y caminos	11. Existencia de infraestructura ISSEA
1	El Terremoto	0.3175	0.4586	0.3482	0.3175	0.4586
2	La Luz	0.3175	0.3276	0.4062	0.3175	0.3276
3	Santa Rosa	0.4445	0.3276	0.5222	0.4445	0.3276
4	Licenciado Jesús Terán	0.3810	0.4586	0.2901	0.3810	0.4586
5	Pilotos	0.5080	0.4586	0.5222	0.5080	0.4586
6	Amarillas de Esparza	0.4445	0.3931	0.2901	0.4445	0.3931

Una vez creada la matriz normalizada con los pesos correspondientes el siguiente paso en la técnica TOPSIS encuentra la solución ideal positiva y la solución anti ideal, esta matriz se muestra a continuación en las tablas 26 y 27.

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Tabla 26 Matriz de solución anti ideal positiva 1

Concepto	Localidad	1. Total habitantes	2. Número de viviendas	3. Servicios Electricidad	4. Servicios de Drenaje	5. Servicios de Agua Potable
V+	0.0497	0.0917	0.0261	0.0254	0.0229	0.0261
V-	0.0311	0.0655	0.0145	0.0159	0.0164	0.0145

Tabla 27 Matriz de solución anti ideal positiva 2

Concepto	Localidad	7. Tamaño de localidad	8. Distancia por carretera	9. Distancia lineal	10. Carreteras y caminos
V+	0.0508	0.0459	0.0522	0.0318	0.0459
V-	0.0318	0.0328	0.0290	0.0508	0.0328

Luego obtuvimos la proximidad relativa a la solución absoluta la cual se muestra en la tabla 28.

Tabla 28 Matriz de proximidad relativa a la solución absoluta

Número	Localidad	+SOL	-SOL	RP
1	El Terremoto	0.0326	0.0005	0.0150
2	La Luz	0.0456	0.0008	0.0182
3	Santa Rosa	0.0364	0.0008	0.0213
4	Licenciado Jesús Terán	0.0347	0.0003	0.0079
5	Pilotos	0.0191	0.0006	0.0323
6	Amarillas de Esparza	0.0364	0.0000	0.0000

Finalmente, en la Tabla 29 se muestran los resultados obtenidos en el último paso de la aplicación del TOPSIS, que fue la creación de la lista de clasificación de soluciones y se define a la comunidad de Pilotos en el municipio de Asientos como la ubicación ideal para establecer la Jurisdicción sanitaria 4.

CAPÍTULO VI PROPUESTA DE REUBICACIÓN DE JURISDICCIONES

Tabla 29 Lista de clasificación de soluciones

Número	Localidad	Clasificación de la solución
1	El Terremoto	4
2	La Luz	3
3	Santa Rosa	2
4	Licenciado Jesús Terán	5
5	Pilotos	1
6	Amarillas de Esparza	6

Este resultado fue enviado y sometido en la publicación Handbook of Security and Privacy of AI-Enabled Healthcare Systems and Internet of Medical Things, como el capítulo titulado “AIoMT on Intelligent Selection of Location and Clustering Medical Units”

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VII PRODUCTOS Y ACTIVIDADES DERIVADOS DE LA INVESTIGACIÓN

Capítulo VII. Productos y Actividades derivados de la investigación

7.1. Convenio de colaboración UAA-ISSEA.

Se estableció el convenio de colaboración entre la Dirección de Planeación y Desarrollo del Instituto de Servicios de Salud del Estado de Aguascalientes y el estudiante de doctorado, Pablo Rodríguez de León, mediante el cual se comparte la información generada en consulta por el sistema de Expediente Clínico Electrónico de 89 centros de Salud y 5 Hospitales generales para el desarrollo de una herramienta metaheurística híbrida de aprendizaje automático para identificar los factores de riesgo asociados con la patología de Diabetes Mellitus en el estado de Aguascalientes, así como otros puntos importantes, este convenio tiene una vigencia del 1 de enero del 2019 al 1 de enero del 2022, el convenio de colaboración está integrado para su consulta en el Anexo 1.

7.2. Producto de selección de subconjuntos de características de testores típicos.

Este resultado fue presentado en el artículo “Mecanismo híbrido de clasificación para diabetes mellitus, en la población de Aguascalientes, México” en Congreso Mexicano de Inteligencia Artificial 2020.

La primera técnica experimentada fue de selección de subconjuntos de características basada en testores típicos para el cual se llevaron a cabo los siguientes pasos:

1. Con la discretización de los datos de las consultas otorgadas a pacientes de las tres clases establecidas, se consiguió la matriz de aprendizaje.
2. Se creó la Matriz de diferencias.
3. Se creó la Matriz básica.
4. Ya obtenida la matriz básica se creó el total de testores y se reconoció solamente a los testores típicos.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VII PRODUCTOS Y ACTIVIDADES DERIVADOS DE LA INVESTIGACIÓN

Dependencia de datos, sin límite de clases y características a comprar, son algunas de las ventajas que tienen los testores típicos, además no solo especifican las variables de importancia estos también especifican individualmente el peso informacional de las variables importantes.

Mediante la implementación de testores típicos relacionados a una matriz de entrenamiento obtenida a partir de un conjunto de datos conformados con información histórica de pacientes con la patología de diabetes mellitus y sin diabetes, se analizaron 25 variables presentadas en la Tabla 9.

Se emplearon 390 registros de pacientes, repartidos en tres grupos de 130 pacientes cada uno, un grupo de pacientes con diabetes, otro grupo con pacientes con prediabetes y otro grupo de pacientes sin diabetes.

Toda la información de las variables se procesó por medio de testores típicos, que fue desarrollada en RadStudio 10.3.2 en lenguaje Delphi y C++, se generaron cinco testores típicos presentados en la Tabla 30.

Tabla 30 Testores típicos producidos

Testor 1	Sexo	Glucemia	Edo. civil	Índice de mc	Complicaciones.	Tensión Alta
Peso informacional.	40%	80%	100%	60%	100%	40%
Testor 2	Glucemia	Estado civil	Imc	Comp.		
Peso informacional.	80%	100%	60%	100%		
Testor 3	Anteced. Fam. Diab.	Estado civil	Comp.	Tensión Alta		
Peso informacional.	40%	100%	100%	40%		
Testor 4	Sexo	Glucemia	Estado civil	Índice de mc	Complicaciones.	
Peso informacional.	40%	80%	100%	60%	100%	

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VII PRODUCTOS Y ACTIVIDADES DERIVADOS DE LA INVESTIGACIÓN

Testor 5	Ant. Fam.				
	Diab.	Glucemia	Edo civil	Comp.	Edad
Peso informacional.	40%	80%	100%	100%	40%

7.3. Establecimiento de parámetros para redes neuronales

Para cada testor obtenido se diseñó y puso en marcha una RNA densamente interconectada de una capa oculta, considerando como neuronas de entrada cada una de las variables del testor típico, donde se implementó la ecuación 28, propuesta por Piedra Fernández (Piedra Fernández, 2008) para calcular el número de neuronas de la capa oculta, este autor un par de soluciones una formada por una capa oculta y otra de 2 capas ocultas.

$$h = n - m$$

Ecuación 24

Donde

h = Num de neuronas capa oculta.

n = Num de neuronas entrada.

m = Num de neuronas salida.

Estas RNA se implementaron con Python 3.0 empleando librerías de Keras y Tensorflow, se establecieron como medidas de procesamiento 100 pacientes de cada una de las clases diabéticos, prediabéticos y no diabéticos.

A las neuronas de la capa de entrada y de la capa oculta se les implementó activación relu y en las neuronas de la capa de salida se implementó activación softmax, se procesó por 90 épocas en lotes de 15, en la etapa de validación se utilizaron 30 pacientes de cada una de las 3 clases, debido a que la selección de subconjuntos de variables implica el procesamiento todo el conjunto de datos de las tres clases existentes. Las configuraciones topológicas y resultado de precisión de prueba de cada una de las redes neuronales desarrolladas se muestran en la Tabla 31.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VII PRODUCTOS Y ACTIVIDADES DERIVADOS DE LA INVESTIGACIÓN

Tabla 31 configuración de RNA's para testores típicos

Testor	Variables	Nodos Entrada	Nodos Oculta	Nodos Salida	Precisión de Test
1	6	6	5	1	0.5
2	4	4	3	1	0.883
3	5	5	4	1	0.883
4	5	5	4	1	0.883
5	5	5	4	1	0.916

Examinando las precisiones de los cinco clasificadores implementados, se observa que el mejor es el de la red neuronal artificial que procesó las variables del testor típico 5, con el cual se obtuvo un poco más del 91% de precisión de la prueba, determinando a las variables de la Tabla 32, estado civil, existencia de complicación, nivel de glucemia, la edad y antecedentes familiares con diabetes, en subconjunto fueron las variables de importancia para clasificar diabetes mellitus en la población analizada.

Tabla 32. Pesos informacionales del testor resultante de mejor precisión

Variable	Descripción	Peso Informacional.
1	Antecedentes familiares con diabetes	40%
2	Nivel de glucemia	80%
3	Estado civil	100%
4	Existencia de complicación	100%
5	Edad	40%

Una variable importante con un peso informacional de los más altos fue el estado civil, en esta variable se observa que el 48% de los pacientes con diabetes pertenecen al estado casado, que únicamente el 22% son del estado soltero y el restante 30 % pertenecen al resto de los estados.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VII PRODUCTOS Y ACTIVIDADES DERIVADOS DE LA INVESTIGACIÓN

En relación con la variable nivel de glucemia se observa que el 83% pertenece a los pacientes diabéticos que pertenecen a la clase con riesgo y solo un 17% de la población con diabetes son de la clase sin riesgo.

Respecto con la edad se observa una alta incidencia de pacientes con diabetes en los rangos de edades de adulto que es del 61% y senecto que es del 28%, 11% restante pertenece a los demás rangos de edad.

La variable de existencia de complicación la cual fue otra variable con el 100% de peso informacional, presenta que 10% de los pacientes diabéticos son de categoría sin complicación y lamentablemente el 90% de los pacientes con diabetes son de la categoría con complicación.

La otra la variable que se analizó fue antecedentes de familia con diabetes y esta presento que el 67% de los pacientes diabéticos son de la categoría con antecedentes familiares y solo el 33% son pacientes con diabetes y pertenecientes a la categoría sin antecedentes.

7.4. Producto clasificador Ingenuo de Bayes

Paralelamente se experimentó con la técnica de clasificador ingenuo de bayes, esto con el fin de comparar los resultados de precisión de esta técnica con los resultados de implementar clasificación por medio de redes neuronales artificiales.

Se experimentó usando el total de cinco variables determinadas en el testor típico de mayor precisión obtenido en los clasificadores de redes neuronales artificiales, este clasificador se entrenó con un insumo de 100 pacientes confirmados con diabetes mellitus y 100 pacientes no diabéticos, a continuación, se probó el clasificador desarrollado con 30 pacientes diabéticos y 30 pacientes no diabéticos como se muestra en la Tabla 33.

CAPÍTULO VII PRODUCTOS Y ACTIVIDADES DERIVADOS DE LA INVESTIGACIÓN

Tabla 33 Matriz de resultados total.

Diabetes	Si	No.	Total
Si	26	4	30
No	2	18	20

En la tabla 34 se presenta la matriz de resultados generada y se obtuvo una precisión del 88%.

Tabla 34 Matriz de resultados en porcentajes.

Diabetes	Si	No.	Total
Si	86.67%	13.3%	100%
No	10%	90%	100%
Precisión de 88%			

Este resultado fue presentado en el congreso de investigación en el posgrado 2019 UAA con el título de “Algoritmo metaheurístico para clasificación identificación y clasificación de factores importantes de diabetes mellitus en Aguascalientes”

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VIII DISCUSIÓN Y CONCLUSIONES

Capítulo VIII. Discusión y conclusiones

En este capítulo, se presentan las conclusiones con respecto de los objetivos de la investigación, así como la discusión de los resultados obtenidos.

Se logró el objetivo general y específico de diseñar, implementar y poner a punto un algoritmo metaheurístico híbrido de aprendizaje automático para identificar las causas de riesgo asociados con la patología de Diabetes Mellitus en el estado de Aguascalientes, ya que se determinó exitosamente el subconjunto de características y su peso informacional con el que se puede identificar un paciente diagnosticado con la patología de diabetes mellitus mediante CIE10 en la población de Aguascalientes. Se diseñó, puso en marcha y a punto un clasificador fundamentado en redes neuronales artificiales profundas con el cual obtuvo un 91.6% de precisión.

Podemos concluir que no se puede reemplazar o excluir las variables existencia de complicación y estado civil, debido a que estas variables son las de mejor peso informacional obtenido el cual fue del 100%, igualmente se confirma la relevancia de edad, glucemia y antecedente familiar con diabetes que en conjunto con complicación y estado civil forman parte del testor encontrado, el cual toma en cuenta la correspondencia entre la totalidad de variables con las cuales se identifica pacientes con diabetes mellitus.

Se acepta la hipótesis de que mediante el algoritmo metaheurístico híbrido se identificaron los factores en los diversos grupos de pacientes con diabetes mellitus clasificados por las claves del CIE10. Además, se obtuvo exitosamente el objetivo de identificar las causas de desarrollar complicaciones como enfermedad cardíaca, derrame cerebral, ceguera y otros problemas de los ojos, enfermedad renal, neuropatía y otras, las cuales están clasificadas en el CIE 10 en pacientes diagnosticados como diabéticos.

Otro relevante descubrimiento en la población de Aguascalientes, es que la variable de estado civil posee vínculo directo en la determinación de un paciente con diabetes mellitus, ya que es una variable de alta importancia con un 100% de peso informacional y en la clase casado posee una alta concurrencia de pacientes diabéticos y una menor concurrencia de pacientes solteros, esto probablemente por los niveles altos de estrés y el alto nivel de sedentarismo por el cual pasan los pacientes de la clase casados.

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VIII DISCUSIÓN Y CONCLUSIONES

Respecto a la variable antecedentes de familia con diabetes y complicaciones podemos concluir que tienen una fuerte relación debido a que cuando analizamos su combinación de categorías de con complicación y cuenta con antecedentes de familia con diabetes 63 de un total de 65 pacientes resultaron con diagnóstico de diabetes mellitus, en comparación de la combinación de sin complicación y sin antecedentes de familiar con diabetes mellitus de los cuales solo 5 pacientes de un total de 45 resultaron con diagnóstico de diabetes mellitus.

A su vez podemos descartar la hipótesis de que la agrupación de solo las variables de género, glucemia, IMC y tensión arterial son en la población de Aguascalientes, los únicos factores concluyentes de la patología de diabetes mellitus.

Mediante la integración de herramientas como K-means y técnicas de selección multicriterio aplicadas en el sector salud, los centros de salud reunirán todos y cada uno de los factores que implican retos a partir de una logística complicada en tiempo y forma, además de diversos procesos con diferentes demandas, con diferentes niveles de complejidad de protocolos, para funcionar de manera organizada y eficiente, con miras a mejorar la atención en términos de calidad, detección temprana de patologías como diabetes mellitus así como la reducción de costos.

Podemos concluir que el algoritmo K-means modificado es una herramienta sólida que puede ser condicionada a diferentes condiciones de distribución y clusterización. En esta investigación se adaptó para trabajar con centroides fijos y móviles debido a la naturaleza del problema, y se obtuvo un resultado muy positivo, además se comprueba la hipótesis de que existen focos geográficos de concentración de pacientes con la patología de diabetes mellitus a través de la obtención y análisis de totales de pacientes determinados con la patología de diabetes mellitus.

Es importante mencionar que la inclusión de una nueva jurisdicción es factible, ya que la distancia entre la jurisdicción sanitaria y sus centros de salud se reducirá al menos a la mitad, y se atenderán más de 65,000 pacientes, más de 3000 pacientes con patología de diabetes mellitus confirmada y más de 3000 pacientes con prediabetes, lo cual es importante, ya que la diabetes mellitus es la 9ª causa de muerte a nivel mundial en 2019 según La Organización Mundial de la Salud (World Health Organization, 2021) y la 2ª causa de muerte en México de enero a agosto de 2020 (99,733 muertes), con una tasa nacional de 7.8 y una tasa estatal de 5,8 y una tasa estatal de 5,5.8 y una tasa estatal de 5.2 defunciones por cada 10,000 habitantes, y 2ª causa en la mortalidad general en Aguascalientes en 2018 con el 14.67% del total de defunciones según el Gobierno del Estado de Aguascalientes (Gobierno del Estado de Aguascalientes, 2019).

CAPÍTULO VIII DISCUSIÓN Y CONCLUSIONES

Al incluir la jurisdicción sanitaria 4, se atenderán 30,595 pacientes de la jurisdicción sanitaria 1 y 34,973 de la jurisdicción sanitaria 2, reduciendo la carga de trabajo de las jurisdicciones sanitarias 1 y 2 en 5.5% y 27%, respectivamente.

En el análisis de las posibles localidades para establecer la jurisdicción sanitaria 4, se concluye que no existe una regla de restricción en cuanto a la posición de la oficina jurisdiccional, sin embargo, se encontró que la conformación de una jurisdicción sanitaria está determinada por la regionalización, con el fin de planificar la prestación de servicios, proporcionar el uso óptimo de los recursos y responder a las necesidades particulares de atención médica (Ejecutivo & Estado, 2011) por lo que la técnica TOPSIS fue determinante para seleccionar la localidad.

Se puede concluir que las ponderaciones para cada característica en TOPSIS fueron dadas por el experto en planificación urbana y a través del asesoramiento del experto en planificación urbana, se estableció un nivel de importancia para las características analizadas por TOPSIS:

1. Total, de habitantes, esta característica tiene una dinámica y depende de dos factores: El natural y el social, que en realidad son la natalidad y la mortalidad, así como la emigración. El tamaño de la población tiene cuatro factores primordiales: natalidad, mortalidad, inmigración y emigración. Además, se puede precisar un grupo de factores suplementarios de la población, como la proporción de sexos, la distribución por edades, el patrón de distribución espacial, etc. El tamaño de la localidad viene determinado por la relación entre la población y el tamaño del territorio, comúnmente esta variable está vinculada a las mejores oportunidades de empleo y acceso a los servicios, y al número de viviendas.
2. Caminos y carreteras, suma de distancias lineales y viales, en este caso de distancias basadas en el principio de centralidad de la teoría de Christaller (Valbuena, 2013) en la que se distribuyen y clasifican espacios públicos que brindan algunos servicios a los pobladores de un área determinada.
3. La población con servicios de electricidad, drenaje, agua potable y salud, ya que solo en conjunto y no por separado determinan la consolidación de la localidad (Cuesta, 1985), estas características están comúnmente ligadas a la dependencia del tamaño de la población y el tamaño de la comunidad y la dispersión territorial, características que se consideran más importantes (Secretaría Distrital de Desarrollo Económico, 2010).

Se concluye que dadas las características de la localidad de Pilotos, municipio de Asientos, es la localidad ideal, ya que a diferencia de las otras posibles localidades, cuenta con una población mayor a 1000 personas, cuenta con los servicios básicos (luz, agua, drenaje, teléfono) necesarios para el funcionamiento de una jurisdicción, cuenta con acceso vial a través de la carretera federal 43 y

TESIS TESIS TESIS TESIS TESIS

CAPÍTULO VIII DISCUSIÓN Y CONCLUSIONES

además cuenta con un centro de salud actualmente en funcionamiento, con las instalaciones para poder establecer la jurisdicción sanitaria 4.

Es importante mencionar que esta investigación es una propuesta basada en datos disponibles y de libre acceso, ya que no es posible acceder a información sobre gastos de operación y presupuestos de los centros de salud y jurisdicciones sanitarias.

Para futuros trabajos y para experimentar con diferentes resultados, más de un centroide variable e integrar y posicionar geográficamente otras nuevas jurisdicciones en el estado, además de la comparación de resultados con otras técnicas de búsqueda exhaustiva como el Modelo de Mezcla Gaussiana o las técnicas de Agrupación Espacial Basada en Densidad de Aplicaciones con Ruido (DBSCAN).

Se tiene pensado a futuro añadir este prototipo de clasificadores como una herramienta de sugerencias y auxiliar de diagnóstico, aplicándolo en diversas patologías y apoyando a los médicos en la consulta del expediente clínico electrónico del ISSEA.

CAPÍTULO IX REFERENCIAS

Capítulo IX Referencias.

- Barrios Arce, J. I. (2020). Inteligencia Artificial y Machine Learning para todos. Retrieved from Juan Barrios Tecnología y Educación website: https://www.juanbarrios.com/inteligencia-artificial-y-machine-learning-para-todos/#Aprendizaje_vs_Inteligencia
- Blesa, M. J., Blum, C., Roli, A., & Sampels, M. (2005). Hybrid Metaheuristics: Second International Workshop. *LNCS*, 3636.
- Blum, C., & Roli, A. (2004). Proceedings of the First International Workshop on Hybrid Metaheuristics. *International Workshop on Hybrid Metaheuristics*.
- Cifuentes Ramos, F. A. (2016). Clasificación automática de Tweets utilizando K-NN y K-Means como algoritmos de clasificación automática, aplicando TF-IDF y TF-RFL para las ponderaciones. *Pontificia Universidad Católica de Valparaíso Facultad de Ingeniería Escuela de Ingeniería Informática*, 0(0), 0.
- Cuesta, G. F. (1985). Problemas teóricos sobre la definición de unidades espaciales de análisis. *Ería*, pp. 213–225. Retrieved from [http://www.unioviedo.net/reunido/index.php/RCG/article/download/907/837%5Cnhttps://onedrive.live.com/edit.aspx/Documentos2/Doctorado?cid=14f67dacc40b137&id=documents?&wd=target\(Papers/Er?a.one%7CD72C4EE8-18E7-45CF-8314-7A4FBA3714E3/Problemas teóricos sob](http://www.unioviedo.net/reunido/index.php/RCG/article/download/907/837%5Cnhttps://onedrive.live.com/edit.aspx/Documentos2/Doctorado?cid=14f67dacc40b137&id=documents?&wd=target(Papers/Er?a.one%7CD72C4EE8-18E7-45CF-8314-7A4FBA3714E3/Problemas%20te%C3%B3ricos%20sob)
- Ejecutivo, P., & Estado, D. E. L. (2011). Manual de organización tipo de jurisdicción sanitaria. *Gaceta Del Gobierno*., (308), 40.
- FID, F. I. (2015). Factores de Riesgo asociados a diabetes mellitus. Retrieved from

CAPÍTULO IX REFERENCIAS

- Factores de Riesgo asociados a diabetes mellitus website:
<http://www.idf.org/node/26455?language=en>
- Gennaro, J. D. (1983). *The New Physical Fitness: Exercise for Everybody 1983*. Englewood, Colorado: Morton Publishing Company.
- Gobierno del Estado de Aguascalientes. (2019). *Informe de Gobierno del Estado de Aguascalientes 2019*. Aguascalientes.
- Gonzalez Gonzalez, I. (2022). *APLICACIÓN DE INTELIGENCIA AMBIENTAL PARA LA PLATAFORMA DE CÁMARAS INTELIGENTES INTEOX*. UNIVERSIDAD COMPLUTENSE DE MADRID.
- INEGI, I. N. de. (2014). *ESTADÍSTICAS A PROPÓSITO DEL DÍA MUNDIAL DE LA DIABETES*. Retrieved from <https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2014/poblacion4.pdf>
- Ishizaka, A., & Nemery, P. (2013). Topsis. In *Multi-Criteria Decision Analysis: Methods and Software* (Vol. 1, pp. 213–221). <https://doi.org/10.1002/9781118644898.ch8>
- Kolios, A., Mytilinou, V., Lozano-Minguez, E., & Salonitis, K. (2016). A comparative study of multiple-criteria decision-making methods under stochastic inputs. *Energies*, 9(7), 1–21. <https://doi.org/10.3390/en9070566>
- Mamani, Z. (2015). *Aplicación de la minería de datos distribuida usando algoritmo de clustering K-means para mejorar la calidad de servicios de las organizaciones modernas*. 150.
- Mansilla A, M. E. (2010). ETAPAS DEL DESARROLLO HUMANO. *Revista de Investigación En Psicología*, Vol.3(No.2), 115.

CAPÍTULO IX REFERENCIAS

- MasedaTarin, M. (2019). *Reducción de la dimensionalidad mediante métodos de selección de características en microarrays de ADN* (Universitat Oberta de Catalunya). Retrieved from <https://openaccess.uoc.edu/bitstream/10609/89668/11/mmasedaTFG0119memoria.pdf>
- Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones. Historia*, 55. Retrieved from <ftp://decsai.ugr.es/pub/usuarios/castro/Material-Redes-Neuronales/Libros/matich-redesneuronales.pdf>
- Mondelo, P. R., Enrique Torada, G., Comas Uriz, S., Castejón Vilella, E., & Bartolomé Lacambra, E. (1999). Confort y Estrés Térmico. In *Ergonomía 2* (p. 19,20). España: Alfaomega.
- MU. (2013). Factores de riesgo para la diabetes tipo 2. Retrieved from Centro Médico de La Universidad de Maryland website: <http://umm.edu/Health/Medical/SpanishEncy/Articles/Factores-de-riesgo-para-la-diabetes-tipo-2#ixzz3cTC6YmIR>
- Narendra, P. M., & Fukunaga, K. (1977). A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, 26, 917–922.
- Olson, D. L. (2004). Comparison of weights in TOPSIS models. *Mathematical and Computer Modelling*, 40(7–8), 721–727. <https://doi.org/10.1016/j.mcm.2004.10.003>
- ORGANIZACIÓN PANAMERICANA DE LA SALUD. (2015). Clasificación Internacional de las Enfermedades. In *Organizacion panamericana de la salud* (3ra ed.). Retrieved from http://www.paho.org/hq/index.php?option=com_content&view=article&id=3

CAPÍTULO IX REFERENCIAS

- 561%3Aclasificacion-internacional-enfermedades-cie&catid=2641%3Acha-clasificacion-internacional-enfermedades-cie&Itemid=2560&lang=en
- Parra Martínez, C. (2014, December). Planificación trimestral del entrenamiento para un paciente con hipertensión arterial. *EFDeportes.Com Revista Digital*, 1. Retrieved from <https://www.efdeportes.com/efd199/entrenamiento-con-hipertension-arterial.htm>
- Piedra Fernández, J. A. (2008). *Aplicación de los sistemas neurodifusos a la interpretación automática de imágenes de satélite* (Universidad de Almeria). <https://doi.org/978-84-8240-784-5>
- Raidl, R. (2006). A Unified View on Hybrid Metaheuristics. *Hybrid Metaheuristics (LNCS 4030)*, 1–12. https://doi.org/10.1007/11890584_1
- Rodríguez Bocca, P. (2008). Metaheurísticas y Optimización sobre Redes: Introducción y Fundamentos. In *Facultad de Ingeniería Universidad de la Republica INRIA*. Rennes, Francia.
- Salcedo López, F. H. (2015). *Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC*. 184. Retrieved from [http://dspace.unl.edu.ec/jspui/bitstream/123456789/10462/1/ChambaJiménez%2C Sairy Fernanda.pdf](http://dspace.unl.edu.ec/jspui/bitstream/123456789/10462/1/ChambaJiménez%2C%20Sairy%20Fernanda.pdf)
- Sánchez, D. G., & Lazo, C. M. (2002). Modificaciones al algoritmo BT para mejorar sus tiempos de ejecución. *V 20*, 129–136.
- Santiesteban, A., & Pons, P. A. (2003). Lex: Un Nuevo Algoritmo para el Cálculo de los Testores Típicos. *Revista Ciencias Matemáticas V 21*, 88–95.
- Secretaria Distrital de Desarrollo Económico. (2010). *Diagnóstico: ALTERNATIVAS DE LA POBLACIÓN RURAL EN POBREZA PARA GENERAR INGRESOS*

CAPÍTULO IX REFERENCIAS

SOSTENIBLES. 1, 49.

Serrano, A. J., Soria, E., & Martin, J. D. (2010). *Redes Neuronales Artificiales.*

Valencia, España: Universitat de València.

Shulcloper, J. R., Alba, C., & Lazo, C. (1995a). Introducción a la teoría de testores típicos. In *Serie Verde* (Vol. 50). Ciudad de México.

Shulcloper, J. R., Alba, C., & Lazo, C. (1995b). Introducción al reconocimiento de Patrones: Enfoque Lógico Combinatorio. *Serie Verde, 51.*

Torres, J. (2018). *Deep Learning — Aprendizaje profundo Introducción práctica con Keras (primera parte)* (Primera ed; What this space, Ed.). Barcelona, España.

Torres, M. ., Ponce, E. ., Ochoa, C. ., Torres, A., & Díaz, E. (2009). *Mecanismos de Aceleración en Selección de Subconjuntos de Características Basada en el Peso Informacional de las Variables para Aprendizaje no Supervisado.* Orlando, Florida.

Torres, M. D. (2010). *Metaheurísticas Híbridas en Selección de Subconjuntos de Características para Aprendizaje no Supervisado.* Universidad Autónoma de Aguascalientes.

Torres Soto, M. D. (2010). *Metaheurísticas Híbridas en Selección de Subconjuntos de Características para Aprendizaje no Supervisado.* Universidad Autónoma de Aguascalientes.

Triantaphyllou, E. (2000). *Multi-Criteria Decision Making Methods.* 5–21.

https://doi.org/10.1007/978-1-4757-3157-6_2

Valbuena, L. G. B. (2013). Microeconomics approaches in the christaller's central places theory. *Ensayos Sobre Política Económica, 31(70), 67–120.*

World Health Organization. (2021). Diabetes Notas Descriptivas. Retrieved from

CAPÍTULO IX REFERENCIAS

notas descriptivas website: <https://www.who.int/es/news-room/fact-sheets/detail/diabetes>

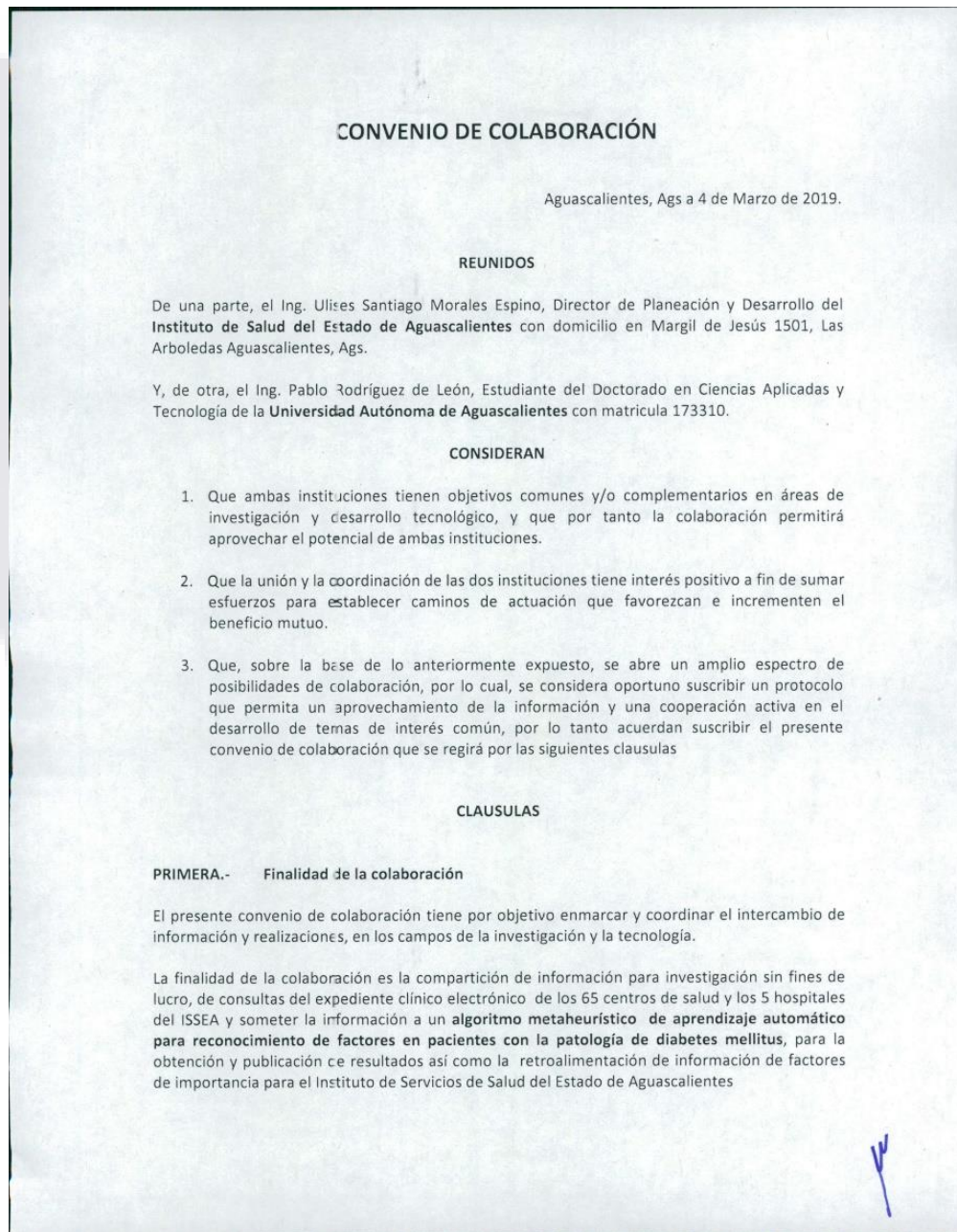
Yoon, K. P., & Hwang, C.-L. (1995). Multiple Attribute Decision Making. In *Methods for Qualitative Data* (p. 23).



CAPÍTULO X ANEXO

Capítulo X Anexo.

10.1 Convenio de colaboración



CAPÍTULO X ANEXO

SEGUNDA.- Formalización

La información proporcionada por el ISSEA estará libre de datos y diagnósticos sensibles o comprometedores de identificación de pacientes así como datos contables o financieros del Expediente Clínico Electrónico.

Los datos obtenidos de los 65 centros de salud y los 5 hospitales serán concentrados en un repositorio o base de datos con datos específicos de consultas de pacientes con diagnósticos relacionados con la patología de Diabetes Mellitus.

TERCERA.- Relación de recursos humanos y materiales

Cada una de las instituciones aportará recursos humanos y materiales al convenio para que se pueda llegar a buen fin.

CUARTA.- Equipo humano

Por parte de la UAA, la persona responsable del resguardo de la información proporcionada será el MITC. Pablo Rodríguez de León alumno del Doctorado en Ciencias Aplicadas y Tecnología.

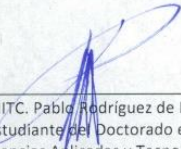
QUINTA.- Duración

El presente convenio de colaboración es vigente desde el día 1 de Enero de 2019 y tendrá una vigencia de 4 años, prorrogables previo acuerdo escrito de las partes.

Y, en prueba de conformidad, ambas partes firman el presente convenio de colaboración, en la ciudad y en la fecha mencionada.



 Ing. Ulises Santiago Morales Espino
 Director de Planeación y Desarrollo
 ISSEA



 MITC. Pablo Rodríguez de León
 Estudiante del Doctorado en
 Ciencias Aplicadas y Tecnología
 UAA