



**UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO**  
**DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA**  
**INFORMACIÓN**



**MINERÍA DE DATOS APLICADA EN UN SISTEMA DE APOYO A LA**  
**TOMA DE DECISIONES MÉDICAS**

TESIS PARA OBTENER EL GRADO DE:

**MAESTRO EN ADMINISTRACIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN**

PRESENTA:

**L.I.A. ISMAEL ZÁRATE HERNÁNDEZ**

BAJO LA DIRECCIÓN DE:

**DR. GUILLERMO DE LOS SANTOS TORRES**

EN CODIRECCIÓN:

**DR. ERIC RAMOS MÉNDEZ**

CUNDUACÁN, TABASCO, A: MAYO 2025



**UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO**  
**DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA**  
**INFORMACIÓN**



**MINERÍA DE DATOS APLICADA EN UN SISTEMA DE APOYO A LA**  
**TOMA DE DECISIONES MÉDICAS**

TESIS PARA OBTENER EL GRADO DE:

**MAESTRO EN ADMINISTRACIÓN DE TECNOLOGÍAS DE LA INFORMACIÓN**

PRESENTA:

**L.I.A. ISMAEL ZÁRATE HERNÁNDEZ**

BAJO LA DIRECCIÓN DE:

**DR. GUILLERMO DE LOS SANTOS TORRES**

EN CODIRECCIÓN:

**DR. ERIC RAMOS MÉNDEZ**

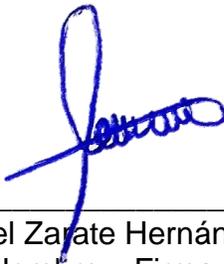
CUNDUACÁN, TABASCO, A: MAYO 2025

## Declaración de Autoría y Originalidad

En la Ciudad de Cunduacán, Tabasco, el día 19 del mes Mayo del año 2025, el que suscribe **Ismael Zarate Hernández** alumno del Programa de **Maestría en Administración de Tecnologías de la Información** con número de matrícula **222h19011**, adscrito a la **División Académica de Ciencias y Tecnologías de la Información**, de la Universidad Juárez Autónoma de Tabasco, como autor de la Tesis presentada para la obtención del grado de Maestría y titulada **Minería de Datos aplicada en un Sistema de Apoyo a la toma de decisiones Médicas** dirigida por el Dr. Guillermo de los Santos Torres y el Dr. Eric Ramos Méndez

**DECLARO QUE:** El trabajo recepcional es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la LEY FEDERAL DEL DERECHO DE AUTOR (Decreto por el que se reforman y adicionan diversas disposiciones de la Ley Federal del Derecho de Autor del 01 de Julio de 2020 regularizando y aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita. Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad o contenido de la Tesis presentada de conformidad con el ordenamiento jurídico vigente.

Cunduacán, Tabasco a 19 de Mayo de 2025.



---

Ismael Zarate Hernández  
Nombre y Firma

Cunduacán, Tabasco a 07 de mayo de 2025  
Oficio No. 0776/2025/DACYTI/D

Asunto: Autorización de impresión de Tesis

**C. Ismael Zarate Hernández**

Egresado de la Maestría en Administración de Tecnologías de la Información

En virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudios de Posgrado vigente en la Universidad, informo a Usted que se autoriza la impresión del trabajo recepcional "**Minería de datos aplicada en un sistema de apoyo a la toma de decisiones médicas**", para presentar examen y obtener el Grado de Maestro en Administración de Tecnologías de la Información.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

Atentamente

  
**M.T.E. Óscar Alberto González González**  
Director

  
UNIVERSIDAD JUAREZ  
AUTÓNOMA DE TABASCO  
DIVISIÓN ACADÉMICA DE  
CIENCIAS Y TECNOLOGÍAS  
DE LA INFORMACIÓN

C.c.p. Dr. Eddy Arquímedes García Alcocer. - Encargado del Despacho de la Coordinación de Posgrado DACYTI  
Archivo.  
Consecutivo.

M.T.E. OAGG/EAGA 

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.  
Cunduacán, Tabasco, México.  
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870  
E-mail: direccion.dacyti@ujat.mx



**UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO**

**DIVISIÓN ACADÉMICA DE CIENCIAS  
Y TECNOLOGÍAS DE LA INFORMACIÓN**



**F8: Cesión de Derechos**

**Carta de Cesión de Derechos**

Cunduacán, Tabasco, a 15 de mayo de 2025.

Por medio de la presente manifestamos haber colaborado como AUTORES en la producción, creación y/o realización de la obra denominada **“MINERÍA DE DATOS APLICADA EN UN SISTEMA DE APOYO A LA TOMA DE DECISIONES MÉDICAS”**.

Con fundamento en el artículo 83 de la Ley Federal del Derecho de Autor y toda vez que, la creación y/o realización de la obra antes mencionada se realizó bajo la comisión de la Universidad Juárez Autónoma de Tabasco; entendemos y aceptamos el alcance del artículo en mención, de que tenemos el derecho al reconocimiento como autores de la obra, la Universidad Juárez Autónoma de Tabasco mantendrá en un 100% la titularidad de los derechos patrimoniales por un periodo de 20 años sobre la obra en la que colaboramos, por lo anterior, cedemos el derecho patrimonial exclusivo en favor de la Universidad.

**COLABORADORES**

\_\_\_\_\_  
**Ismael Zarate Hernández**  
Alumno

**TESTIGOS**

\_\_\_\_\_  
**Dr. Guillermo de los Santos Torres**

\_\_\_\_\_  
**Dr. Eric Ramos Méndez**

## Dedicatorias

### **A mi esposa:**

A mi querida esposa Nancy Ivette, por estar siempre a mi lado querida esposa, por creer en mí y confiar en que siempre saldremos adelante a pesar de las muchas adversidades que se han presentado en los años que llevamos juntos como esposos.

Gracias por tus palabras de aliento, tus consejos que fueron de gran ayuda y de motivación en los momentos cuando pensaba que ya no podía continuar. Este triunfo es de los dos.

### **A mis queridos hijos:**

Gael y Ximena los quiero mucho, ustedes son el motivo por el que siempre deseo superarme y demostrarles que se puede lograr grandes cosas con mucho sacrificio y trabajo duro.

## **Agradecimientos**

En primer lugar, expreso mi agradecimiento a Dios todo poderoso por darme la vida y permitir que hoy cumpla con una de las metas que me propuse en la vida. Doy gracias porque a pesar de las adversidades y obstáculos que siempre se presentaron, Él me permitió salir adelante.

Al Consejo Nacional de Humanidades Ciencia y Tecnología (CONAHCYT) por haberme otorgado la beca que permitió cubrir todos los gastos que realicé durante mis estudios de maestría.

A mis directores de tesis, al Dr. Guillermo de los Santos Torres y al Dr. Eric Ramos Méndez, por el apoyo que me han brindado, su paciencia y sus consejos, los cuales fueron de gran ayuda para concluir con mi tema de investigación y por sus palabras de aliento para seguir adelante y la confianza que tuvieron en mí.

Agradezco de forma muy especial a mi compañero Héctor Eduardo Vidal de la Cruz por sus buenos consejos y el apoyo con sus conocimientos que siempre estuvo dispuesto a compartir conmigo, para la realización de esta tesis, a mis demás compañeros con los cuales conviví dentro y fuera de las aulas, gracias por su apoyo y amistad.

Gracias a la Universidad Juárez Autónoma de Tabasco (UJAT), por aceptarme y abrirme las puertas para cumplir con una de mis metas, estudiar y obtener el grado de Maestro de Administración en Tecnología de la Información (MATI).

A todos, mis agradecimientos y respetos.

# Índice de Contenido

Índice de Tablas .....	xii
Índice de Figuras .....	xiii
Capítulo 1. Introducción .....	1
1.1 Planteamiento del problema .....	1
1.2 Preguntas de investigación .....	5
1.3 Hipótesis o supuestos .....	5
1.4 Objetivos .....	5
1.4.1 Objetivo general .....	5
1.4.2 Objetivos específicos .....	5
1.5 Justificación .....	5
1.5.1 Alcances del proyecto .....	7
1.6 Metodología .....	8
1.6.1 Enfoque de la investigación .....	8
1.6.2 Fuentes de investigación .....	8
1.6.3 Técnicas de recolección de datos .....	8
1.6.4 Desarrollo del proceso KDD .....	8
1.6.5 Metodología aplicada al sistema de recomendaciones médicas .....	9
Capítulo 2. Marco Teórico .....	10
2.1 Minería de datos .....	10
2.2 Metodologías de minería de datos .....	11
2.2.1 Proceso KDD .....	11
2.2.2 Metodología SEMMA .....	12
2.2.3 Metodología CRISP-DM .....	13
2.3 Técnicas de minería de datos predictivos más utilizados en la actualidad: .....	15
2.3.1 Árboles de decisión .....	15
2.3.2 Regresión logística .....	16
2.3.3 Redes Neuronales Artificiales .....	16

2.3.4	Redes bayesianas.....	17
2.4	Un Sistema de Soporte a la Toma de Decisión (SSD).....	17
2.5	Marco referencial .....	19
2.5.1	La minería de datos en apoyo a la toma de decisiones clínicas .....	19
2.5.2	Aplicación y exploración de Big Data Mining en Medicina Clínica .....	19
2.5.3	La aplicación de métodos de Minería de Datos para el proceso de diagnóstico de enfermedades .....	20
2.5.4	Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas	21
2.5.5	Minería predictiva de datos en medicina clínica: temas y directrices actuales	22
2.5.6	Aplicación de Minería de Datos para el pronóstico de la evolución de la diabetes en México .....	23
2.6	Marco conceptual.....	24
2.6.1	Diagnóstico erróneo .....	24
2.6.2	Tratamiento médico .....	24
2.6.3	Segundo nivel de atención médica .....	24
2.6.4	Prescripción médica.....	24
2.6.5	Minería de datos .....	24
2.6.6	Metodología KDD.....	25
2.6.7	Sistema de Soporte a la Toma de Decisión .....	25
2.7	Marco legal .....	25
2.7.1	Ley Federal de Protección de Datos Personales en Posesión de los Particulares. ....	25
2.7.2	Ley General de Salud. ....	25
2.7.3	Norma Oficial Mexicana NOM-024-SSA3-2012, Sistemas de información de registro electrónico para la salud. Intercambio de información en salud. ....	26
2.7.4	Licencia de Software libre GPL.....	26

Capítulo 3.	Aplicación de la metodología y desarrollo .....	27
3.1	Fase 1: Obtención de los datos .....	27
3.2	Fase 2: Diseño del <i>dataset</i> .....	28
3.3	Fase 3: Evaluación de tres técnicas de minería de datos .....	33
3.3.1	Técnicas de agrupamiento .....	33
3.3.2	Técnicas de reglas de asociación .....	37
3.3.3	Técnicas de árboles de decisión .....	41
Capítulo 4.	Resultados.....	47
4.1	Código escrito para desarrollar el modelo .....	48
4.1.1	Cargar y preprocesar los datos (primera parte) .....	48
4.1.2	Definir funciones para cargar datos y modelos (segunda parte) .....	49
4.1.3	Configurar la interfaz de usuario con Streamlit .....	50
4.1.4	Predecir medicamentos y mostrar resultados .....	51
4.2	Interfaz gráfica del sistema de recomendaciones médicas.....	52
4.3	Evaluaciones del sistema de recomendaciones médicas .....	55
4.3.1	Evaluación de funcionalidad del sistema. ....	55
4.3.2	Comentarios finales del médico.....	56
4.3.3	Evaluación de la usabilidad del sistema.....	56
4.3.4	Otros resultados.....	57
Capítulo 5.	Conclusiones, recomendaciones y trabajos futuros .....	58
5.1	Conclusiones .....	58
5.2	Recomendaciones .....	59
5.3	Trabajos futuros .....	59
Referencias	.....	61
Glosario	.....	66
Apéndice A	Artículo publicado en la Revista Investigación Aplicada, un Enfoque en la Tecnología.....	67

Apéndice B Artículo publicado en la International Association for Digital Transformation and Technological Innovation ..... 68

Apéndice C Artículo publicado en la International Conference on Information Technology & System ..... 69

Anexo A. Formato de evaluación del sistema de recomendaciones médicas ..... 70

Universidad Juárez Autónoma de Tabasco.  
México.

## Índice de Tablas

<b>Tabla 1</b> Quejas recibidas por la CONAMED en el año 2022.....	3
<b>Tabla 2</b> Motivos presentados en las quejas según causa detallada (tratamiento médico) .....	3
<b>Tabla 3</b> Centroides localizados .....	36
<b>Tabla 4</b> Resultados de la evaluación de las tres técnicas de minería de datos .....	47

## Índice de Figuras

<b>Figura 1</b> Proceso corto de minería de datos .....	11
<b>Figura 2</b> Descripción general de los pasos que constituyen el proceso KDD .....	12
<b>Figura 3</b> Iteración de fases en la metodología SEMMA .....	13
<b>Figura 4</b> Fases de la metodología CRISP-DM .....	14
<b>Figura 5</b> Componentes que forman un SSD .....	18
<b>Figura 6</b> Metodología implementada .....	27
<b>Figura 7</b> Portal de datos abiertos de México .....	28
<b>Figura 8</b> Tabla de URGENCIAS .....	29
<b>Figura 9</b> Código para la selección de la tabla urgencias .....	30
<b>Figura 10</b> Código Python para seleccionar los registros del municipio de Tacotalpa ...	30
<b>Figura 11</b> Procedimiento para generar el archivo csv .....	31
<b>Figura 12</b> Procedimiento para la combinación de las tablas .....	32
<b>Figura 13</b> Dataset utilizado en el proyecto .....	33
<b>Figura 14</b> Código implementado para obtener la gráfica de codos .....	34
<b>Figura 15</b> Gráfica de codo .....	35
<b>Figura 16</b> Código implementado para la obtención de los centroides .....	35
<b>Figura 17</b> Gráfica de la distribución de los cluster y centroides identificados .....	37
<b>Figura 18</b> Distribución del número de ítems por transacción .....	38
<b>Figura 19</b> Número de transacciones .....	38
<b>Figura 20</b> Top 5 de Afecciones principales .....	40
<b>Figura 21</b> Código escrito para identificar las reglas de asociación generadas .....	41
<b>Figura 22</b> Código para dividir el conjunto de datos en características y variable objetivo .....	42
<b>Figura 23</b> Código para dividir el conjunto de datos en características y etiquetas de entrenamiento .....	42
<b>Figura 24</b> Código utilizado para entrenar el clasificador del árbol de decisión .....	43
<b>Figura 25</b> Estructura del árbol .....	44

<b>Figura 26</b> Cuadro de predictores y su importancia del modelo.....	44
<b>Figura 27</b> Cuadro de predictores y la importancia del modelo.....	45
<b>Figura 28</b> Seudocódigo que muestra la primera parte del código escrito para este proyecto.....	49
<b>Figura 29</b> Seudocódigo que muestra la segunda parte del código escrito para este proyecto.....	50
<b>Figura 30</b> Seudocódigo que muestra la tercera parte del código escrito para este proyecto .....	51
<b>Figura 31</b> Seudocódigo que muestra la cuarta parte del código escrito para este proyecto .....	52
<b>Figura 32</b> Interfaz gráfica del sistema de recomendaciones médicas .....	54
<b>Figura 33</b> Carta de aceptación del artículo Uso de la minería de datos en un sistema de apoyo a la toma de decisiones para determinar el tratamiento clínico adecuado.....	67
<b>Figura 34</b> Carta de aceptación del artículo Modelo de minería de datos para la detección de enfermedades en pacientes de primer nivel de atención médica .....	68
<b>Figura 35</b> Carta de aceptación del artículo Development of a data mining model using the clustering technique and the A priori algorithm to determine medical treatments in second level medical care hospitals .....	69

## **Título.**

Minería de Datos aplicada en un Sistema de Apoyo a la Toma de Decisiones Médicas.

## **Resumen**

Las complicaciones causadas por los errores en la prescripción de tratamientos médicos son un problema que pueden ocurrir en cualquiera de los niveles de atención de salud. Para los pacientes, una mala prescripción puede ocasionar desde reacciones alérgicas y toxicidad hasta la ineficacia del tratamiento, lo que puede agravar su estado de salud e incluso poner en riesgo su vida. Para los médicos los errores en la prescripción pueden afectar su credibilidad y provocar consecuencias legales o disciplinarias, especialmente si se demuestra que hubo negligencia o falta de actualización en sus conocimientos médicos. Esta investigación se realizó con la finalidad de desarrollar un modelo de minería de datos para implementarlo en un sistema de apoyo a la toma de decisiones en el área médica. Este sistema fue diseñado con la finalidad de asistir a los médicos del segundo nivel de atención médica en la determinación de tratamientos clínicos adecuados para los pacientes, con el propósito de minimizar el margen de error en la prescripción médica.

**Palabras claves:** Minería de datos, algoritmo *k-means*, *Clustering*, Errores de diagnósticos.

**Title.**

Data Mining applied in a Medical Decision Support System

**Abstract.**

Complications caused by errors in the prescription of medical treatments are a problem that can occur at any level of health care. For patients, poor prescribing can result in everything from allergic reactions and toxicity to ineffective treatment, which can aggravate their health and even put their lives at risk. For physicians, prescribing errors can affect their credibility and lead to legal or disciplinary consequences, especially if it is proven that there was negligence or lack of updating of their medical knowledge. The objective of this research was to develop a data mining model for the implementation of a clinical decision support system. This system was designed with the aim of assisting physicians at the second level of medical care in determining appropriate clinical treatments for patients, in order to help reduce the margin of error in medical prescribing.

**Keywords:** Data mining, k-means algorithm, Cluster, Clustering, Diagnostic errors.

## Capítulo 1. Introducción

### 1.1 Planteamiento del problema

En el 2019 la Organización Mundial de la Salud (OMS) menciona que al año son 134 millones de pacientes los que sufren algún tipo de daño que son causados por una atención poco segura, lo que ocasiona 2.3 millones de muertes tan solo en países con bajos y medianos ingresos. Cuatro de cada 10 pacientes sufren daños en la atención primaria de salud y los servicios ambulatorios (Organización Mundial de la Salud [OMS], 2019).

En el año 1991 la Universidad de Harvard realizó una investigación donde se analizaron 30,195 registros hospitalarios seleccionados al azar, en este análisis se identificaron que 1,133 pacientes sufrieron lesiones causadas por un mal tratamiento médico. (Brennan, *et al*,1991).

Los errores de medicación en servicios de urgencias son variados, pero se destacan principalmente los errores de prescripción, que pueden incluir dosis incorrectas, frecuencia de administración errónea y omisiones. Un estudio reveló que hasta el 54% de los errores de medicación ocurren en la fase de prescripción, lo que subraya la importancia de esta etapa en el proceso de atención (ISMP Canada Safety Bulletin, 2020).

Los errores derivados de una inadecuada prescripción médica en los servicios hospitalarios representan un problema significativo que pueden provocar graves consecuencias para los pacientes y para las instituciones prestadoras de servicios de salud. Estos errores no solo afectan la seguridad de los pacientes, sino también pueden generar costos adicionales para los sistemas de salud y para los familiares de los pacientes.

Los errores en la prescripción de medicamentos pueden generar consecuencias graves para el estado de salud de un paciente, incluyendo reacciones desfavorables y la ineficacia del tratamiento recomendado. Los errores derivados de una mala medicación no solo ponen en riesgo el bienestar físico del paciente, sino que también pueden

ocasionar una progresión no deseada de la enfermedad o la iniciación de nuevas complicaciones (Smith, *et al*, 2021).

Los errores de medicación también generan importantes costos económicos adicionales tanto para los pacientes como para los hospitales. McCarthy *et al* (2020) Mencionan que la aplicación de un tratamiento médico inadecuado puede ocasionar graves impactos económicos para los pacientes, ya los familiares de los pacientes pueden tener gastos adicionales debido a tratamientos correctivos y estancias más prolongadas de un paciente.

En México, El 08 de septiembre del año 2017 el Diario Oficial de la Federación dio a conocer un documento titulado Acciones Esenciales para la Seguridad del Paciente donde se menciona que el 8% de los pacientes que son hospitalizados pueden sufrir algún tipo de daño durante su estancia hospitalaria y el 2% muere, los más destacado de este informe es que el 62% de los incidentes que ocurren durante la atención médica son prevenibles, por lo que se debe enfocar en ellos para ofrecer una atención médica más segura

En México los datos que actualmente se tienen sobre las quejas relacionadas con los tratamientos médicos están registrados en su mayoría en los informes emitidos por la Comisión Nacional de Arbitraje Médico (CONAMED) y en el estado de Tabasco la Comisión Estatal de Conciliación y Arbitraje Médico de Tabasco (CECAMET). Las cifras que esta dependencia de gobierno tiene registradas son aquellas que reciben por parte de los usuarios que interponen una queja, pero existe la posibilidad de que las cifras sean mayores a las registradas ya que no todos los usuarios del servicio médico que tuvieron problemas provocados por una mala atención médica o por algún tipo de error médico interpusieron su queja.

La CONAMED, en su Anuario estadístico 2022, identificó que en el período comprendido del 01 de enero al 31 de diciembre del año 2022 se presentaron 3,042 quejas sobre el servicio de atención médica, de las cuales 1,028 están relacionadas con los problemas relacionados con los tratamientos médicos, como se puede observar en la

Tabla 1, las inconformidades por parte de los pacientes que ocupan el primer lugar son las que están relacionadas con los tratamientos médicos (CONAMED, 2023).

**Tabla 1**

*Quejas recibidas por la CONAMED en el año 2022*

<b>Motivos de la inconformidad</b>	<b>Quejas concluidas</b>
Tratamiento médico	1042
Diagnóstico	1028
Relación médico paciente	503
Tratamiento quirúrgico	305
Auxiliares de diagnóstico	142
Atención del parto y puerperio	9
Deficiencias administrativas y/o institucionales	7
Accidentes e incidentes	6
<b>Total</b>	<b>3,042</b>

*Nota.* Elaboración propia con datos de la CONAMED (2023).

En la Tabla 2 se presentan los principales sub-motivos de las quejas relacionados con las quejas atendidas por la CONAMED y que están relacionadas con los tratamientos médicos.

**Tabla 2**

*Motivos presentados en las quejas según causa detallada (tratamiento médico)*

<b>Motivos por causa detallada</b>	<b>Total</b>
Tratamiento inadecuado	395
Diferimiento	282
Complicaciones secundarias: desinformación	169
Secuelas	131
Negación del servicio	52
Exceso terapéutico	4
<b>Tratamiento médico</b>	<b>1,042</b>

*Nota.* Elaboración propia con datos de la CONAMED (2023).

De acuerdo con los datos presentados en la tabla 2 se puede observar que, dentro de la categoría de problemas con los tratamientos médicos, el tratamiento inadecuado ocupa el primer lugar de las quejas recibidas en la CONAMED en el año 2022.

La CECAMET menciona en su Informe Anual de actividades del año 2022 recibieron en total 52 quejas, de estas quejas 90% estaban relacionadas con el tratamiento clínico que los pacientes recibieron (CECAMET, 2023).

De acuerdo con la información que la CONAMED publicó en el año 2023, los errores de diagnóstico constituyen un serio problema de salud pública y estos deben ser considerados como una amenaza potencial para la seguridad de los pacientes, ya que al realizar un diagnóstico equivocado los pacientes están expuestos a recibir un tratamiento que, en lugar de ayudar a recuperar su malestar físico, puede provocar daños más severos a su estado de salud y en el peor de los casos la muerte.

Ante estos datos, se puede afirmar que el uso de la tecnología puede representar un papel crucial en la prevención de errores de medicación y disminuir la aparición de los errores médicos. La minería de datos es una herramienta que puede adoptarse en el área médica para analizar la información que se genera en este campo y con base a este análisis, crear medidas que favorezcan y apoyen a los profesionales de la salud al momento de realizar la toma de decisiones cuando emitan un tratamiento médico.

El sistema de recomendaciones médicas que será el resultado de esta investigación contribuiría a reducir significativamente la incidencia de errores diagnósticos, sino que también mejorará la calidad y oportunidad de que los pacientes reciban el tratamiento médico adecuado, fortaleciendo así la eficiencia del sistema de salud en su conjunto.

## 1.2 Preguntas de investigación

- ¿Las variables seleccionadas para crear el *dataset*, pueden ser útiles para entrenar un modelo de minería de datos que permita determinar un tratamiento médico?
- ¿Qué técnica de minería de datos muestra un mejor rendimiento para desarrollar un modelo de minería de datos que permita determinar un tratamiento médico: técnicas de agrupamiento, árboles de decisión o reglas de asociación?

## 1.3 Hipótesis o supuestos

El sistema de recomendaciones médicas genera sugerencias de medicamentos que coinciden en gran medida con las prescripciones realizadas por un médico general en un entorno clínico de primer y segundo nivel.

La evaluación realizada por un médico general sugiere que el sistema tiene un nivel aceptable de fiabilidad, permitiendo su potencial implementación en entornos reales de atención médica.

## 1.4 Objetivos

### 1.4.1 Objetivo general

Desarrollar un sistema de apoyo a la toma de decisiones clínicas basado en minería de datos que permita identificar y recomendar tratamientos médicos.

### 1.4.2 Objetivos específicos

- Diseñar el *dataset* para entrenar el modelo de minería de datos.
- Probar las técnicas de árboles de decisión, reglas de asociación y agrupamiento para determinar la técnica que sea más adecuada para este proyecto.
- Desarrollar un modelo de minería de datos para implementarlo en un sistema de recomendaciones médicas para la determinación de tratamientos médicos.

## 1.5 Justificación

Independientemente del nivel de desarrollo que los países tengan, los problemas relacionados con problemas que surgen a raíz de un tratamiento inadecuado representan

en la actualidad un serio problema de salud pública, existiendo riesgos en todos los niveles de atención médica.

En tal sentido es necesario que los gobiernos implementen medidas que ayuden a disminuir los problemas causados por errores médicos, una medida podría ser la creación de herramientas tecnológicas para el diagnóstico que ayuden a minimizar el riesgo de cometer un error médico. Como lo mencionan Bélginova et al (2019) el campo médico cuenta con un gran volumen de información médica que requiere un correcto procesamiento y posterior aprovechamiento en el diagnóstico y tratamiento de diversas enfermedades. El adelanto de la tecnología informática ofrece enormes oportunidades para recopilar, procesar, administrar e investigar información médica.

La minería de datos es una herramienta que puede aportar múltiples beneficios al campo de la medicina, en general la minería de datos en la medicina tiene el potencial de realizar diagnósticos más precisos, ayuda a generar tratamientos con alto índice de efectividad y mejora la gestión médica de los pacientes. Si se combinan la información clínica del paciente y la minería de datos, se podrían generar diagnósticos más precisos y esto permitiría prescribir tratamientos más efectivos. También posibilita una gestión más eficaz, eficiente y económica de los recursos sanitarios al identificar riesgos, predecir enfermedades en ciertos segmentos de la población o pronosticar la duración del ingreso hospitalario.

En este trabajo de investigación, mediante el uso de minería de datos, se analizaron los registros contenidos en la base de datos de Urgencias Médicas disponibles en la plataforma de datos abiertos del gobierno de México, con el fin de desarrollar y entrenar un modelo de minería de datos que pueda ser incorporado en un sistema de apoyo a la toma de decisiones médicas, que ayude a determinar el tratamiento adecuado a los pacientes.

Los beneficios que se generarían al implementar el sistema de apoyo a la toma de decisiones serán los siguientes:

- El sistema será diseñado para apoyar a los médicos en la determinación del tratamiento médico más adecuado para cada paciente.
- Prevenir y disminuir la prescripción de tratamientos erróneos.
- Al determinar el tratamiento adecuado, ayudaría ahorrar el desperdicio de recursos médicos (medicamentos).
- Los pacientes recibirían el tratamiento más adecuado.
- Al recibir el tratamiento adecuado, aumentaría la confianza de los pacientes en el servicio médico que reciben en los hospitales ubicados en la región sierra del estado de Tabasco y esto evitará que acudan a servicios de salud privada para una nueva valoración médica que ocasionaría gastos adicionales para los pacientes.

## 1.6 Alcances del proyecto

El objetivo final de este proyecto es desarrollar un sistema de apoyo a la toma de decisiones médicas usando técnica de minería de datos, que ayude a determinar tratamientos clínicos adecuados. El desarrollo del proyecto comprende las siguientes etapas:

### **Diseño del *Dataset*:**

- Recolección de los datos para generar el *dataset*.
- Preprocesamiento y limpieza de los datos (eliminación de valores nulos, codificación, normalización).
- Diseñar el *dataset* con el que será entrenado el modelo de minería de datos.

### **Desarrollo del modelo de minería de datos:**

- Se seleccionó una técnica de minería de datos para la generación del modelo, considerando entre árboles de decisión, reglas de asociación y técnicas de agrupamiento.

**Implementar el sistema de apoyo a la toma de decisiones clínicas mediante la integración del modelo de minería de datos desarrollado:**

- Implementación del sistema mediante el uso del lenguaje de programación Python.

#### **Pruebas y validación del sistema:**

- Realización de pruebas con datos de prueba para evaluar la precisión del sistema.
- Ajustes finales al modelo y al sistema según los resultados obtenidos.

## **1.7 Metodología**

### **1.7.1 Enfoque de la investigación**

El enfoque de investigación utilizado para el desarrollo de este proyecto fue cuantitativo basado en el diseño no experimental, esto de acuerdo con lo que menciona Monje-Álvarez (2011), la investigación cuantitativa, es un proceso sistemático y ordenado que sigue una serie de pasos definidos.

### **1.7.2 Fuentes de investigación**

Para el desarrollo de este proyecto se recurrió a fuentes de información primarias y fuentes secundarias.

Como fuente primaria se empleó la entrevista que se realizó con el personal médico de la institución donde se llevó a cabo la investigación.

Las fuentes secundarias que se utilizaron son las siguientes: tesis de maestría y doctorado, artículos científicos, libros sobre minería de datos y artículos publicados sobre el tema de errores de diagnóstico.

### **1.7.3 Técnicas de recolección de datos**

Los datos que utilizados en la investigación fueron recolectados de la página de datos abiertos del gobierno de México, descargados en el sitio web <https://datos.gob.mx/busca/dataset/urgencias>.

### **1.7.4 Desarrollo del proceso KDD**

En la etapa tres de la metodología de esta investigación, se desarrolló el modelo de minería de datos, para generar el modelo se utilizó el proceso *KDD*, este proceso se aplicó a las tres técnicas de minería de datos que se manejaron en este proyecto, con la finalidad de determinar cuál de las tres técnicas generaría resultados más acertados al realizar el análisis del *dataset* creado para el desarrollo de esta investigación. El proceso

fue determinante para elegir las variables que se utilizaron para diseñar el *dataset* con el que el modelo de minería de datos fue entrenado.

### **1.7.5 Metodología aplicada al sistema de recomendaciones médicas**

Para desarrollar el modelo y el sistema denominado “Sistema de Recomendaciones Médicas” se utilizó la metodología *waterfall* o cascada la cual consta de seis pasos que consisten en: requisitos, diseño, implementación, pruebas, desarrollo y mantenimiento.

Al seguir cada uno de estos procesos se desarrolló el sistema de recomendaciones médicas. En el capítulo cuatro de esta investigación se muestran los resultados que se obtuvieron al aplicar la metodología *waterfall*.

México.

Autónoma de Tabasco.

## Capítulo 2. Marco Teórico

### 2.1 Minería de datos

De acuerdo con Daza-Vergaray (2016) se puede definir la minería de datos como el proceso de analizar y procesar un volumen grande de datos de diferentes fuentes de alojamiento conocidas como base de datos para obtener información que sea relevante para las empresas y pueda utilizarse en beneficio de estas, mayormente para un beneficio económico.

Actualmente con el avance de la tecnología se genera una gran cantidad de datos dentro de las empresas, pero estos datos sin el procesamiento y análisis adecuado no tienen ninguna relevancia para las empresas, es aquí donde radica la importancia de la minería de datos ya que, gracias al análisis que se hace de los datos, se pueden obtener resultados que proporcionen información valiosa para que las empresas puedan tomar decisiones, planificar estrategias, realizar la predicción de comportamientos futuros y muchos otros casos de uso.

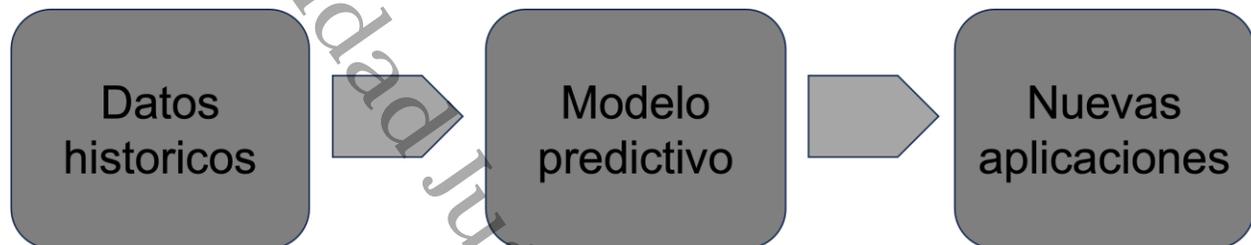
Ahlemeyer y Coleman (2014) mencionan que las decisiones basadas en medidas objetivas son mejores que las decisiones basadas en opiniones subjetivas que pueden ser engañosas y sesgadas. La capacidad de extraer conocimiento útil, pero generalmente oculto, de los datos es cada vez más importante en el mundo competitivo de hoy en día. Cuando los datos se utilizan para la predicción, el comportamiento futuro del negocio es menos incierto y eso solo puede ser una ventaja.

La minería de datos tiene sus antecedentes en varias disciplinas relacionadas con el análisis de datos y la inteligencia artificial. Algunos de los antecedentes importantes de la minería de datos son: Inteligencia Artificial (IA), Aprendizaje automático (*Machine Learning*), Estadística, Bases de datos y Reconocimiento de patrones. La evolución que han tenido cada una de estas disciplinas y tecnologías se han establecidos las bases para la minería de datos como una disciplina ayude a realizar el análisis de grandes volúmenes de datos para extraer conocimiento que sea útil.

La minería de datos utiliza datos del pasado (datos históricos de una organización) para predecir las actividades de los futuros solicitantes. La figura 1 muestra cómo trabaja la minería de datos.

**Figura 1**

Proceso corto de minería de datos



Nota. Elaboración propia basada en Ahlemeyer-Stubbe y Coleman (2014).

## 2.2 Metodologías de minería de datos

Para el modelado de minería de datos es necesario utilizar algunas metodologías, las más utilizadas son: el proceso *KDD, Sampling, Exploring, Modifying, Modeling, and Assessing (SEMMA)* y *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Cada una de estas metodologías tienen características diferentes que las hacen útiles para diversos casos de aplicación.

A continuación, se describen cada una de estas metodologías:

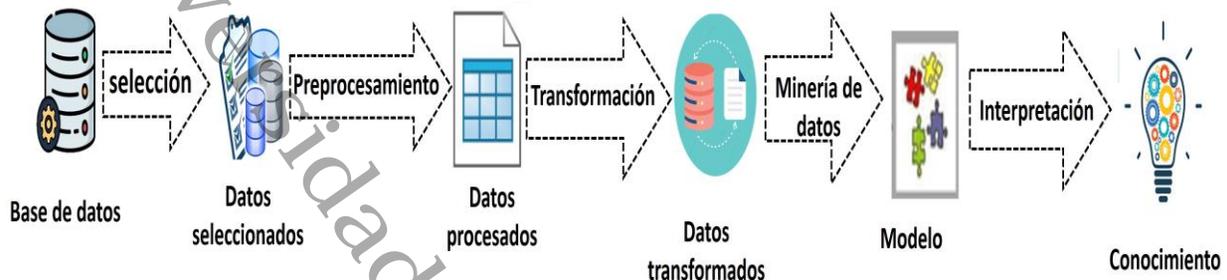
### 2.2.1 Proceso KDD

Cestero y Caballero (2018) mencionan que los datos por sí solos no tienen mucho valor, al igual que el oro o la plata sirven de poco cuando están formando parte de la acantita o la silvanita. Para que un conjunto de datos tenga un valor relevante es necesario realizar un proceso que sea capaz de transformar esos datos en información y conocimiento que permita tomar decisiones. A este proceso de transformación de datos se le conoce como Extracción de Conocimiento en Base de Datos (*KDD*).

La metodología *KDD* fue propuesta por Fayyad en el año 1996, en esta metodología se proponen cinco fases que consisten en: selección, procesamiento, transformación, minería de datos, evaluación e implementación. Cada uno de estos pasos se muestran en la figura 2.

**Figura 2**

Descripción general de los pasos que constituyen el proceso KDD



*Nota.* Elaboración propia basada en datos de Fayyad et al. (1996, p.29).

### 2.2.2 Metodología SEMMA

La metodología SEMMA (por sus siglas en inglés) es un proceso diseñado para la selección, exploración y modelado para realizar el análisis de grandes volúmenes de datos. Fue propuesta por el SAS Institute Inc. en el año 2012 y se compone de cinco etapas fundamentales, que son las siguientes:

**Muestreo (*Sampling*):** Se toma la muestra de un conjunto de datos, este debe ser lo suficientemente grande para contener información relevante que sea útil al proyecto a desarrollar.

**Exploración (*Exploring*):** se busca conocer a profundidad los datos que se disponen, para realizar este procedimiento se utilizan técnicas de visualización, análisis de agrupamiento y herramientas estadísticas para identificar relaciones, tendencias y patrones relevantes que aporten valor al análisis.

**Modificación (*Modifying*):** en esta etapa de la metodología se preparan los datos para la etapa de modelado. Para realizar este proceso es necesario, seleccionar o construir variables para estar seguros que los datos se encuentren en un formato óptimo para ser ejecutados con algoritmos de minería de datos.

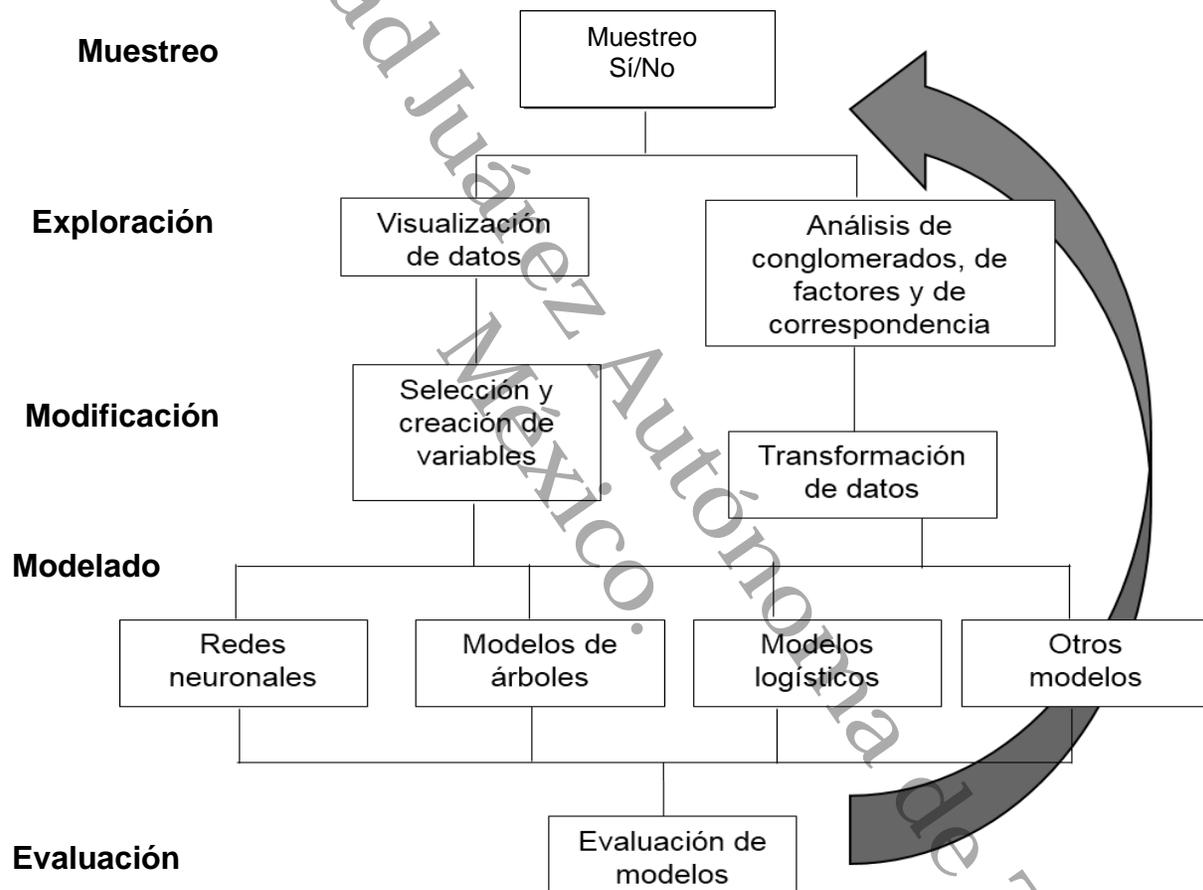
**Modelado (*Modeling*):** En esta etapa se selecciona y aplica la técnica de modelado más adecuada para realizar el análisis del conjunto de datos. Para ejecutar esta fase se crean modelos analíticos utilizando algoritmos específicos con el fin de identificar patrones y generar predicciones de los datos previamente preparados.

**Evaluación (Assessing):** Se evalúa el rendimiento y la validez del modelo seleccionado en la fase anterior, para determinar que si los resultados obtenidos son fiables y confiables.

En la figura 3 se muestra la iteración de las etapas aplicadas en la metodología SEMMA

**Figura 3**

*Iteración de fases en la metodología SEMMA*



*Nota.* Elaboración propia basado en Moine (2013).

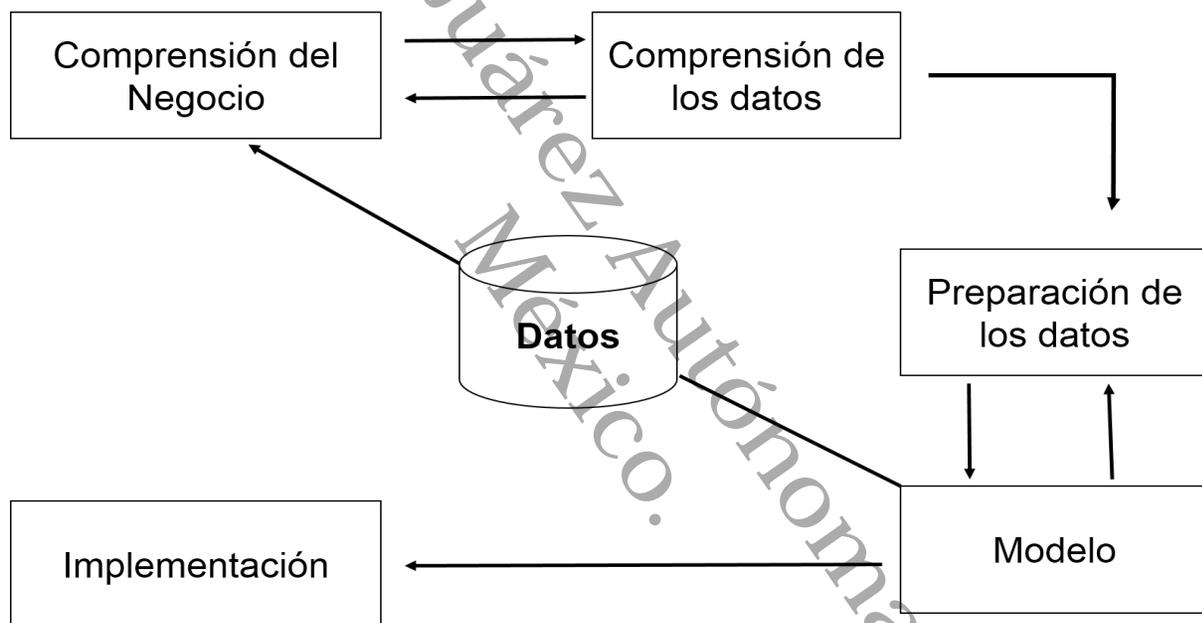
### 2.2.3 Metodología CRISP-DM

La metodología CRISP-DM (por sus siglas en inglés) fue propuesta en el año 1999 por las empresas SPSS, DaimlerChrysler y NCR. Su desarrollo se basó en la experiencia práctica de sus creadores en proyectos reales de minería de datos. Esta metodología

está dividida en cinco fases las cuales consisten en el entendimiento del negocio, el entender los datos, preparar los datos, modelado de los datos, evaluación e implementación del modelo. Dichas fases no tienen que seguirse de manera formal, debido a que presentan tareas generales que, según el nivel de abstracción pueden convertirse en específicas. La figura 4 ilustra el funcionamiento de la metodología *CRISP-DM*.

**Figura 4**

*Fases de la metodología CRISP-DM*



*Nota.* Elaboración propia basado con Radicelli-Garcia, *et al.* (2021).

**Fase de comprensión del negocio:** esta etapa de la metodología es sumamente importante ya que en esta fase se busca entender a profundidad el problema a resolver desde la perspectiva del negocio, para definir de forma clara los objetivos del proyecto, de manera que los resultados obtenidos sean adecuados y útiles para apoyar la toma de decisiones.

Dentro de las actividades a desarrollar en esta fase se consideran: determinar los objetivos del negocio, valorar la situación actual del negocio, dar a conocer los objetivos de la minería de datos y realizar el plan para el desarrollo del proyecto.

**Fase de comprensión de los datos:** la actividad que se realiza en esta fase de esta metodología es la recolección de los datos que serán utilizados para el desarrollo del proyecto, los datos recolectados deben ser de utilidad y estar relacionados.

**Fases de preparación de los datos:** En esta etapa los datos se limpian, seleccionan, transforman y organizan para ser usados en el modelado.

**Fase de modelados:** en esta fase se debe elegir la técnica de modelado más adecuada para el proyecto, para realizar esta etapa de la metodología se deben definir los objetivos del modelo, seleccionar los algoritmos a utilizar, dividir el conjunto de datos y entrenar el modelo con el fin de que el modelo seleccionado cumpla con los objetivos del proyecto y permita interpretar y comprender los resultados.

**Fases de evaluación:** en esta fase se realiza la evaluación del modelo que se propuso anteriormente, considerando criterios de calidad.

**Fase de implementación:** es la transformación en acciones del conocimiento adquirido luego de haber construido y validado el modelo propuesto. Estas acciones se verán aplicadas en el proyecto de minería de datos especificado y deben coincidir con los objetivos planteados en la fase de comprensión del negocio (Radicelli-Garcia, *et al.*, 2021, p. 20-32).

## **2.3 Técnicas de minería de datos predictivos más utilizados en la actualidad:**

### **2.3.1 Árboles de decisión**

Hernández *et al* (2005) señalan que los árboles de decisión son posiblemente uno de los métodos de aprendizaje más sencillo de utilizar y comprender. Un árbol de decisión tiene una estructura organizada de forma jerárquica, esto permite que la decisión final se determine siguiendo las condiciones de todas las jerarquías, desde el inicio del árbol hasta una de sus hojas. Este tipo de modelos son utilizados en diferentes ámbitos como

el médico, legal comercial, estratégico, matemático y lógico, entre otros. En los árboles de decisión las opciones que surgen de una condición son mutuamente excluyentes, lo que permite analizar una situación y siguiendo las condiciones llegar a una sola acción o decisión a tomar (Hernández, *et al.*,2005).

### **2.3.2 Regresión logística**

Este es una extensión de la regresión tradicional, que puede modelar una salida binaria, la cual representa comúnmente la ocurrencia o no de algún evento (James *et al.*, 2013). La regresión lógica se utiliza para modelizar datos en los que la variable objetivo o dependiente es binaria, es decir, la variable dependiente puede tomar el valor 1 con una probabilidad de éxito  $p$ , o es el valor 0 con la probabilidad de fracaso  $1 - p$ . el objetivo principal es desarrollar un modelo de tipo regresión que relacione la variable binaria con las variables independientes.

### **2.3.3 Redes Neuronales Artificiales**

Las Redes Neuronales Artificiales (RNA) son sistemas conexionistas dentro del campo de la Inteligencia Artificial, las cuales, dependiendo del tipo de arquitectura neuronal, puede tener diferentes aplicaciones. Pueden ser utilizadas para el reconocimiento de patrones, la comprensión de información y la reducción de la dimensionalidad, el agrupamiento, la clasificación y la visualización, etc.

Las redes neuronales artificiales solo pueden trabajar de dos modos, el aprendizaje supervisado y el aprendizaje no supervisado:

**Aprendizaje supervisado:** para que la red aprenda se le proporciona un conjunto de datos de entrada sin etiquetarlos para que la red identifique patrones, estructuras o agrupamientos ocultos dentro de los datos. Los datos de entrada son propagados hacia adelante hasta que la activación alcanza la neurona de la capa de salida para comparar la respuesta calculada por la red con la que se desea obtener.

**Aprendizaje no supervisado:** la red es entrenada solo con un conjunto de datos de entrada sin información sobre las respuestas correctas y la red debe auto enseñarse. Casi siempre esta estructura suele deberse a redundancia o agrupamiento en el conjunto

de datos. Este tipo de aprendizaje es útil especialmente para las tareas de agrupamiento y reducción de dimensionalidad (Hernández, *et al.*, 2005).

#### **2.3.4 Redes bayesianas**

De acuerdo con Hernández *et al*, la teoría de la probabilidad y los métodos bayesianos son una de las técnicas que más se han utilizado en problemas de inteligencia artificial. Una de las características principales de la teoría bayesianas es el uso explícito de la teoría de la probabilidad para cuantificar la incertidumbre. La teoría bayesiana permite un doble uso: descriptivo y predictivo. En los modelos descriptivos, las redes bayesianas se enfocan en descubrir de las relaciones de independencia y la relevancia que existe entre cada una de las variables. Esto permite que el modelo resultante represente de forma explícita todas las relaciones de interés, permitiendo una fácil comprensión de la estructura subyacente de los datos.

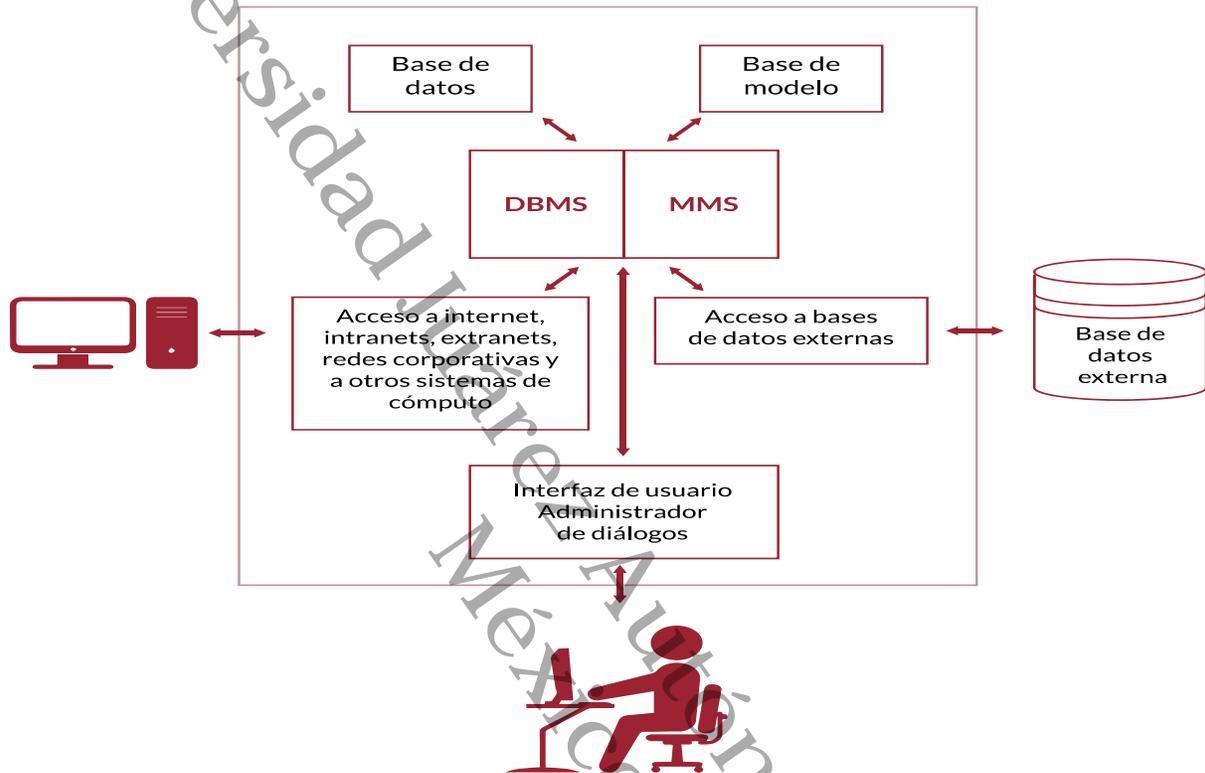
En cuanto al uso predictivo, realiza inferencias a partir de los datos, induciendo modelos probabilísticos que después pueden ser utilizados para razonar sobre nuevos valores observados. Permite calcular de forma explícita la probabilidad asociada a cada una de las hipótesis posibles, lo que constituye una gran ventaja sobre otras técnicas (Hernández, *et al.*, 2005).

### **2.4 Un Sistema de Soporte a la Toma de Decisión (SSD)**

De acuerdo con Stair, *et al* (2010) “los sistemas de apoyo a la toma de decisiones se definen como un conjunto de personas organizadas, procesos, software, base de datos y dispositivos que trabajan en conjunto para respaldar la toma de decisiones que ayuden a resolver problemas específicos (stair, *et al.*, 2010)

De acuerdo con las definiciones anteriores se puede entender que los SSD son aplicaciones de computadoras junto con un componente humano que puede filtrar a través de grandes cantidades de datos y escoger entre numerosas opciones. Al generar resultados basados en el análisis de los datos los SSD permiten que la toma sea más eficiente. En la figura 5 se observa los componentes que forman un SSD.

**Figura 5**  
*Componentes que forman un SSD*



*Nota.* Stair, et al. (2010).

Los SSD son herramientas que ayudan a las personas a tomar decisiones informadas y basadas en datos. Estos sistemas utilizan datos, modelos analíticos y técnicas de visualización para proporcionar información relevante y ayudar a los usuarios a evaluar diferentes alternativas antes de tomar una decisión.

Los SSD que utilizan la minería de datos no necesitan conocimientos previos por parte del decisor. En su lugar, el sistema está diseñado para encontrar patrones y relaciones nuevas e insospechadas en un volumen grande de datos, después de realizar el análisis de los datos el sistema aplica el conocimiento que descubre. Esto resulta muy útil cuando el conocimiento *a priori* es limitado o inexistente (Ozaydin et al., 2016).

## **2.5 Marco referencial**

### **2.5.1 Minería de datos como apoyo en la toma de decisiones clínicas**

Romero-Zaldivar *et al* (2022) realizaron una revisión de la literatura con el fin de identificar, las técnicas más precisas de predicción y su aplicación en la medicina clínica. Las fuentes que utilizaron para su investigación fueron las bases de datos *IEEE Xplore*, *Science Direct*, *ACM Digital Library* y *SciELO*, y los motores de indexación *Google Scholar* y *Scopus* con un rango de fechas desde 2008 a 2019.

Las técnicas de minería de datos que estos autores analizaron fueron, árboles de decisión, regresión logística, redes neuronales artificiales, máquinas de vector de soporte (SVM, por sus siglas en inglés) y redes bayesianas.

Con la revisión literaria que estos investigadores realizaron, identificaron varias áreas de aplicación de las técnicas de la minería predictiva. Una de ellas es el análisis de imágenes médicas para la detección de enfermedades, de las técnicas que analizaron identificaron que los árboles de decisión ofrecen mejores resultados en términos de rendimiento y precisión, seguidos por las redes neuronales artificiales y las máquinas de vectores de soporte.

Romero-Zaldívar revisó investigaciones realizadas sobre la aplicación de técnicas de minería de datos en el pronóstico de la mortalidad en pacientes con cáncer de colon y cáncer de mama, y concluyó que la regresión logística es la técnica que produce los mejores resultados en términos de precisión.

También determinaron que para que un modelo sea efectivo tiene que cumplir con tres criterios importantes los cuales son: la representación del problema, el poder explicativo de su salida y la capacidad de generar conocimiento.

### **2.5.2 Uso y exploración de Big Data Mining en el área de la Medicina Clínica**

La minería de datos puede ser utilizada para apoyar las decisiones clínicas en los diagnósticos clínicos, los médicos a menudo se preocupan por los diversos síntomas de las enfermedades. Una enfermedad puede presentar diferentes síntomas en diferentes pacientes, mientras que muchas enfermedades tienen manifestaciones clínicas iguales o

similares. Esto complica la determinación del diagnóstico de una enfermedad, lo que puede derivar en diagnósticos incorrectos o errores de diagnósticos y en consecuencia provocaría una incorrecta determinación del tratamiento médico (Zhang *et ál.*, 2023).

Estos autores realizaron una revisión de teorías y tecnologías de *big data mining* y sus posibles aplicaciones en la medicina clínica.

Con los datos que obtuvieron llegaron a la conclusión sobre la necesidad de abrir centros analíticos que generalicen la información de diferentes tipos de fuentes sobre la base de la minería de datos. Los investigadores mencionaron la gran necesidad que actualmente existe de formar personal capacitado para trabajar en dichos centros y que puedan analizar grandes volúmenes de datos, interpretar los resultados y formularlos adecuadamente para una comprensión accesible a la mayoría de las formas y preparar soluciones de gestiones basadas en los datos.

De acuerdo con la investigación que estos autores realizaron se determinó que la minería de datos desempeña un papel sumamente importante en el análisis de la información que se genera en el área de la medicina.

### **2.5.3 Aplicación de métodos de Minería de Datos para la determinación de diagnóstico de enfermedades**

Belginova *et al* (2019) mencionan que actualmente en el campo médico se genera una gran cantidad de información que necesita de un adecuado procesamiento para su posterior aprovechamiento en la determinación de los diagnóstico clínicos para establecer el tratamiento médico de múltiples enfermedades. El desarrollo de la tecnología informática brinda grandes oportunidades para recopilar, procesar, administrar e investigar información médica para comprender mejor los complejos procesos biológicos de la vida y ayudar a resolver problemas en el diagnóstico y el tratamiento en las instituciones médicas.

Belginova *et al* (2019) mencionan en su investigación que el análisis inteligente brinda una gran oportunidad para identificar y estudiar patrones que están ocultos en los grandes conjuntos de datos médicos. Los autores sugieren que la mejora en la precisión

de la aplicación *Medical Decision Support* (MDS) se puede lograr mediante el modelado de datos imprecisos y temporales.

El propósito principal de la investigación de estos autores fue verificar la posibilidad principal del procesamiento analítico de los datos disponibles mediante métodos de aprendizaje automático y determinar la precisión con que el modelo de aprendizaje automático puede tener un valor práctico. Realizaron pruebas para la creación de un algoritmo que sea útil en la determinación de un diagnóstico de enfermedades.

Los resultados que obtuvieron fue la presentación de datos en forma de grupos homogéneos (*Clusters*). El algoritmo que propusieron permite calcular la independencia dinámica entre los *Clusters* obtenidos y las interacciones con la base de datos médica.

#### **2.5.4 Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas**

Hernández y Sánchez (2012) realizaron una investigación usando dos modelos matemáticos para determinar de qué manera la minería de datos puede contribuir al diagnóstico de enfermedades, centrándose en la hipertensión arterial (HTA). Para procesar y analizar los datos de esta investigación se implementó la metodología *CRISP-DM* 1.0 y se complementó con la herramienta *WEKA* 3.6.2.

El problema que estos autores plantearon fue el siguiente: en la actualidad, la hipertensión arterial se ha convertido en una de las primeras causas de muertes en el mundo. Un reporte más reciente de la OMS, publicado en septiembre de 2023, reafirma que uno de cada tres adultos a nivel mundial padece de hipertensión, lo que representa una crisis de salud pública continua.

El objetivo de la investigación fue demostrar, que mediante la combinación de dos modelos matemáticos y el uso de técnicas de minería de datos se puede contribuir al diagnóstico de enfermedades. Para efectuar su investigación utilizaron una base de datos que contenía información de personas que pertenecen a la Universidad de las Ciencias Informáticas. Los resultados que obtuvieron con su investigación fueron los siguientes:

- En el 40% de los casos de los pacientes que padecen HTA están entre 45 y 65 años, son de sexo masculino y de raza mestiza.

- El 34% de las personas que sufren de hipertensión arterial presentan antecedentes patológicos familiares de esta enfermedad y sumado a esto, consumen tabaco.
- El 66 % de los casos con HTA fueron asintomáticos.

Al finalizar su investigación obtuvieron las siguientes conclusiones: aunque los algoritmos se propusieron en el sector de la salud, su uso no está restringido a esta área. Como parte de los resultados de esta investigación se desarrollaron dos modelos, el primero basado en el método de clasificación mediante el uso del algoritmo de árboles de decisión J48 y el segundo basado en la técnica de agrupamiento utilizando el algoritmo *K-means*, con la finalidad de encontrar patrones que están ocultos en los datos clínicos de pacientes con hipertensión arterial (HTA). Además, brindó un aporte práctico basado en que el sistema integral para la atención primaria de la salud y un soporte de toma de decisiones para convertirse en un sistema más robusto, permitiendo un rápido análisis de la información para apoyar a la toma de decisiones médicas. Es importante mencionar que los dos modelos conseguidos fueron evaluados con datos reales, comparando los resultados que los modelos obtuvieron con diferentes procedimientos. La evaluación se realizó utilizando las métricas integradas en WEKA y fueron validadas por expertos en HTA.

### **2.5.5 Minería predictiva de datos en medicina clínica: temas y directrices actuales**

La minería de datos ofrece soluciones metodológicas y técnicas para abordar el análisis de datos médicos y la construcción de modelos de predicción. Una gran variedad de estos métodos requiere pautas generales y simples que pueden ayudar a los profesionales en la selección adecuada de herramientas de minería de datos, en la construcción y validación de modelos predictivos en los ambientes clínicos (Bellazzia y Zupanb, 2016).

Bellazzia y Zupanb (2016) realizaron una revisión de trabajos relacionados con la minería de datos predictivas para discutir el alcance y el papel del área de investigación de la minería de datos predictiva y proponer un marco para hacer frente a los problemas

de construcción, evaluación y explotación de modelos de minería de datos en la medicina clínica.

Con los resultados que se obtuvieron en la investigación se destaca que la minería de datos se está convirtiendo en un instrumento indispensable para los investigadores y profesionales en el área de la medicina clínica. Comprender los principales problemas que subyacen a estos métodos y la aplicación de procedimientos estandarizados y acordados es obligatorio para su implementación y la difusión de los resultados.

### **2.5.6 Aplicación de Minería de Datos para el pronóstico de la evolución de la diabetes en México**

Esta investigación se realizó en el año 2019 con el propósito de desarrollar un prototipo de minería de datos que ayudara a realizar proyecciones de las tasas de mortalidad asociada al E11(diabetes mellitus no insulino dependiente). El funcionamiento del prototipo fue validado con información recabada de fuentes oficiales como: el Instituto Nacional de Estadística y Geografía, el Sistema Nacional de Información en Salud, el Consejo Nacional de Población y el Centro Mexicano para la Clasificación de Enfermedades. Para procesar, manipular y extraer la información obtenida se realizó un proceso de minería de datos con la metodología Cross Industry Standard Process for Data Mining (CRISP-DM) (Flores Guerrero, 2019). El autor de esta investigación utilizó la técnica de agrupamiento y el algoritmo k-means debido a que sus resultados son fáciles de interpretar y su implementación computacional es relativamente sencilla.

Con los resultados de esta investigación se determinó que en la zona metropolitana la alcaldía Venustiano Carranza de la Ciudad de México (CDMX) presentó el mayor incremento en su tasa de mortalidad y para la zona de provincia, Orizaba en el estado de Veracruz tendrá el mayor incremento de su mortalidad para el año 2020. El municipio que presenta el mayor decremento en su tasa de mortalidad dentro de la zona metropolitana es la alcaldía de Miguel Hidalgo de la CDMX y en provincia es el municipio de San Pedro Cholula de estado de Puebla.

## **2.6 Marco conceptual**

### **2.6.1 Diagnóstico erróneo**

Son los “daños provocados al paciente por acción u omisión del médico al ejercer su profesión y sin comentarlo intencionalmente, o bien, la conducta clínica incorrecta en la práctica médica, o como consecuencia de la decisión de aplicar un criterio equivocado” (Santamaria-Benhumea, *et al.*, 2014, P.33).

### **2.6.2 Tratamiento médico**

"El tratamiento médico se refiere a la intervención destinada a curar, mitigar, controlar o prevenir una enfermedad o condición, a través del uso de medicamentos, procedimientos quirúrgicos, terapias físicas o psicológicas." (*Manual de la OMS sobre el manejo de enfermedades comunes*, WHO, 2003).

### **2.6.3 Segundo nivel de atención médica**

En este nivel de atención médica se reciben a los pacientes que son enviados desde el primer nivel de atención y son atendidos por especialistas, quienes ofrecen diagnósticos y tratamientos clínicos más específicos y de mayor complejidad (*Niveles de atención en salud: Estrategias y desafíos*, Revista de Salud Pública, 2008).

### **2.6.4 Prescripción médica**

La OMS define a la prescripción médica como el proceso lógico deductivo, que se basa en la información general y objetiva del problema de salud que tiene el paciente.

### **2.6.5 Minería de datos**

Romero Zaldivar *et al* (2022) define la minería de datos como una disciplina que engloba la estadística, la inteligencia artificial y las tecnologías de bases de datos. El propósito de la minería de datos es obtener patrones comprensibles a partir de gran cantidad de datos, a través de técnicas de análisis de datos como el aprendizaje automático y la estadística.

La minería de datos es el proceso de extracción de información útil, patrones ocultos y relaciones significativas de grandes conjuntos de datos que puedan ser utilizados para la toma de decisiones informadas.

### **2.6.6 Metodología KDD**

El proceso KDD, también conocido como descubrimiento de conocimiento en una base de datos, esta metodología es ampliamente utilizada para el desarrollo de proyectos de minería de datos. En la metodología KDD se tiene que realizar varios pasos para encontrar información útil dentro de grandes volúmenes de datos.

Fayyad *et al* (1996) refieren que el proceso El KDD identifica patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles en los datos.

### **2.6.7 Sistema de Soporte a la Toma de Decisión**

Rashidi *et al* (2017) definen a los SSD como sistemas destinados a apoyar a los responsables de la toma de decisiones en problemas semiestructurados que no podían resolverse completamente mediante algoritmos. Se utilizan para modelar el razonamiento humano y el proceso de toma de decisiones; ambos son capaces de aceptar datos de los usuarios, procesarlos y sugerir soluciones cercanas a las presentadas por los expertos humanos.

## **2.7 Marco legal**

### **2.7.1 Ley Federal de Protección de Datos Personales en Posesión de los Particulares.**

Esta ley se publicó en el año 2019 con el fin de proteger los datos que están en posesión de los particulares, con la finalidad de regular su tratamiento legítimo, controlado e informado, a efecto de garantizar la privacidad y el derecho a la autodeterminación informativa de las personas (Ley Federal de Protección de Datos Personales en Posesión de los Particulares, 2010)

### **2.7.2 Ley General de Salud.**

Esta ley se publicó el 7 de febrero de 1984 para regular el derecho a la protección de la salud que tienen las personas sin ninguna distinción de todo el territorio mexicano. Esta ley establece las bases y modalidades para garantizar que todas las personas tengan acceso a los servicios de salud sin ningún tipo de restricción.

### **2.7.3 Norma Oficial Mexicana NOM-024-SSA3-2012, Sistemas de información de registro electrónico para la salud. Intercambio de información en salud.**

El objetivo de esta norma es regular los Sistemas de Información de Registro de expediente Electrónico para la Salud, así como establecer los mecanismos científicos, éticos, tecnológicos y administrativos con el objetivo de que los proveedores de servicios de salud puedan integrar, utilizar, manipular, preservar y almacenar de forma confidencial los datos de los expedientes clínicos de cada paciente.

Todos los prestadores de servicios de salud que manejan Sistema de Información de Registro Electrónico para la Salud y sean parte del Sistema Nacional de Salud en el territorio nacional están obligados a observar esta norma, así como las personas físicas o morales que cuenten con derechos de propiedad, uso, autoría, distribución y/o comercialización de dichos sistemas conforme lo establece esta norma.

### **2.7.4 Licencia de Software libre GPL**

La Licencia Pública General (GPL), es una licencia que permite a los usuarios la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar el software.

Para que un software sea considerado GPL debe cumplir con las cuatro libertades, las cuales indican que el programa deber ser ejecutado libremente, tener acceso al código fuente para ejecutarlo y modificarlo, redistribuir copias y distribuir copias de versiones modificadas (Stallman, 2002, p. 56).

- La libertad 0 indica que el programa puede ser ejecutado libremente con cualquier propósito.
- La libertad 1 permite estudiar el programa y modificarlo si es necesario. Para cumplir esta libertad es necesario que el usuario tenga acceso al código fuente
- La libertad 2 indica que el usuario puede redistribuir libremente el programa.
- La libertad 3 menciona que el usuario puede distribuir copias de versiones modificadas del software original, permitiendo que otros usuarios puedan beneficiarse de las modificaciones realizadas.

### Capítulo 3. Aplicación de la metodología y desarrollo

En este capítulo se detalla la metodología que se aplicó para desarrollar esta investigación, la metodología implementada se dividió en cuatro fases que consistieron en: la obtención de los datos, los cuales fueron descargados del sitio web <https://datos.gob.mx/busca/dataset/urgencias> de datos abiertos del gobierno de México. Procesamiento y limpieza de los datos para diseñar el *dataset* que se utilizó para entrenar el modelo de minería de datos. En la etapa tres se realizaron pruebas con tres técnicas de minería de datos para elegir la más adecuada a esta investigación y posteriormente con la técnica seleccionada se desarrolló el modelo de minería. En la cuarta etapa se desarrolló el sistema de apoyo a la toma de decisiones que permitirá que los médicos determinen tratamientos médicos más precisos, en la figura 6 se puede observar cada una de las fases de la metodología.

#### Figura 6

*Metodología implementada*



*Nota.* Elaboración propia.

#### 3.1 Fase 1: Obtención de los datos

Los datos que se emplearon para desarrollar este proyecto de investigación fueron obtenidos del sitio web <https://datos.gob.mx/busca/dataset/urgencias> de datos abiertos del Gobierno de México, se descargaron las bases de datos de urgencias de los años 2022, 2021 y 2020. Estos datos fueron obtenidos el mes de marzo del año 2023. En la figura 7 se muestra la interfaz gráfica de la página de donde se obtuvieron estos datos.

**Figura 7**

*Portal de datos abiertos de México*



*Nota.* Elaboración propia.

### **3.2 Fase 2: Diseño del *dataset***

Para diseñar el *dataset* primero se analizó la información que contenían las bases que se descargaron en la fase uno de este proyecto, al realizar el análisis se encontraron las siguientes características:

La base de datos del año 2022 contenía un total de 3,872,562.

La base de datos del año 2021 contenía 4,882,162 registros.

La base de datos del año 2020 contenía 5,532,261 registros.

Las bases de datos contenían las tablas de: URGENCIAS, AFECCIONES, PROCEDIMIENTOS Y MEDICAMENTOS, en la figura 8 se puede observar parte de la estructura de la tabla de URGENCIAS.

**Figura 8**

*Tabla de URGENCIAS*

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
ID	CLUES	fechaalta	EDAD	CVEEDAD	SEXO	ENTRESIDENCIA	MUNRESIDENCIA	DERHAB	TIPOURGENCIA	MOTATE	TIPOCAMA	ENVIADOA	MP	AFECPRIN
3816930	TCSSA001064	03/06/2022	71	5	1	27	15	11	2	2	1	4	8	110X
3957793	TCSSA003963	28/11/2022	24	3	2	27	13	11	2	4	3	4	8	R05X
6400859	TCSSA001076	21/10/2022	28	5	2	27	12	0	1	3	1	1	8	Z358
6400860	TCSSA001076	23/10/2022	29	5	2	27	4	0	1	3	3	4	8	Z321
6400861	TCSSA001076	23/10/2022	24	5	2	27	4	11	2	3	3	4	8	O000
6400862	TCSSA001076	23/10/2022	16	5	2	27	4	0	1	3	3	4	8	Z321
6400863	TCSSA001076	23/10/2022	27	5	2	27	4	0	1	2	3	4	8	Z358
6400864	TCSSA001076	23/10/2022	29	5	2	27	4	0	1	3	3	4	8	Z348
6400865	TCSSA001076	23/10/2022	37	5	2	27	4	0	2	3	3	4	8	O000
6400866	TCSSA001076	26/10/2022	26	5	2	27	4	0	2	3	3	4	8	O000
6400867	TCSSA001076	03/10/2022	23	5	2	27	4	0	2	3	3	4	8	O000
6400868	TCSSA001076	03/10/2022	26	5	2	27	4	0	2	3	3	4	8	Z348
6400869	TCSSA001076	23/10/2022	19	5	2	27	4	0	1	3	3	4	8	Z321
6400870	TCSSA001076	03/10/2022	22	5	2	27	4	0	2	3	3	2	8	O000
6400871	TCSSA001076	21/10/2022	20	5	2	27	8	0	1	3	9	1	8	Z321
6400872	TCSSA001076	27/10/2022	17	5	2	27	4	0	2	3	3	4	8	Z359
6400873	TCSSA001076	23/10/2022	38	5	2	27	4	9	1	3	3	4	8	Z321
6400874	TCSSA001076	24/10/2022	27	5	2	27	4	0	1	3	3	4	8	Z321
6400875	TCSSA001076	03/10/2022	25	5	2	27	4	0	2	3	3	4	8	Z348

*Nota.* Elaboración propia basada en la base de datos de Urgencias del Gobierno de México (2023).

Los datos almacenados en las bases de datos descargadas son los registros que se generan en todos los hospitales de segundo nivel de atención medica de México. Para efectos de este proyecto se utilizaron únicamente los registros de los hospitales ubicados en la región sierra del estado de Tabasco, la cual está conformada por los municipios de Teapa, Jalapa y Tacotalpa. Para obtener los registros de los municipios mencionados anteriormente se utilizó como herramienta principal *Python*, esto debido al gran volumen de registros que cada base de datos contenía. Para realizar este proceso se utilizó como clave principal la variable CLUES de la base de datos de Urgencias. En la figura 9 se parte de los datos almacenados en la tabla de URGENCIAS del año 2022.

**Figura 9**

Código para la selección de la tabla urgencias

```
#Tabla Urgencias
Dataframe1=pd.read_csv("/content/drive/MyDrive/2022/URGENCIAS.csv", delimiter='|', header=None, names=col_names)
Dataframe1.head()
```

<ipython-input-11-5c04301c4d67>:2: DtypeWarning: Columns (8) have mixed types. Specify dtype option on import or set Dataframe1=pd.read\_csv("/content/drive/MyDrive/2022/URGENCIAS.csv", delimiter='|', header=None, names=col\_names)

	ID	CLUES	fechaalta	EDAD	CVEEDAD	SEXO	ENTRESIDENCIA	MUNRESIDENCIA	DERHAB	TIPOURGENCIA	...	MP	AFECF
0	1	TSSSA003732	2022-10-25	7	5	1	28	22	11	2	...	8	
1	2	ZSSSA013172	2022-11-08	22	5	2	0	999	0	1	...	8	R
2	3	ZSSSA001016	2022-11-05	10	4	2	32	36	0	1	...	8	

Nota. Elaboración propia.

Para seleccionar los registros de los municipios mencionados anteriormente se implementó el código *Python* que se muestra en la figura 10. Para realizar este procedimiento se tomó como clave principal la variable clubes de cada municipio.

**Figura 10**

Código *Python* para seleccionar los registros del municipio de Tacotalpa

```
[ ] Urgencias=Dataframe1.loc[Dataframe1['CLUES'] == 'TCSSA004301']
print(Urgencias)
```

	ID	CLUES	fechaalta	EDAD	CVEEDAD	SEXO	ENTRESIDENCIA	MUNRESIDENCIA	DERHAB	TIPOURGENCIA	...	MP	AFECPRIN	IRA	PLANEDA
6408200	6408201	TCSSA004301	2022-03-04	11	4	1	27								
6408201	6408202	TCSSA004301	2022-03-08	45	5	2	27								
6408226	6408227	TCSSA004301	2022-12-03	21	5	2	27								
6408259	6408260	TCSSA004301	2022-03-14	26	5	2	27								
6408673	6408674	TCSSA004301	2022-03-24	10	5	1	27								
...	...	...	...	...	...	...	...								
6724702	6724703	TCSSA004301	2022-05-28	9	5	2	27								
6724718	6724719	TCSSA004301	2022-05-30	3	5	2	27								
6724938	6724939	TCSSA004301	2022-06-04	15	5	1	27								
6726042	6726043	TCSSA004301	2022-06-17	4	5	1	27								
6726417	6726418	TCSSA004301	2022-06-27	34	5	1	27								

Nota. Elaboración propia.

Después de obtener los registros con la clave clubes seleccionada, se guardaron en un archivo con extensión *csv* (*comma-separated values*). En la figura 11 se puede observar el código implementado para realizar este procedimiento.

**Figura 11**

*Procedimiento para generar el archivo csv*

```
[ ] Urgencias.to_csv('/content/drive/MyDrive/2022/Urgenciastacotalpa2022.csv', index=False)
```

```
datos1=pd.read_csv("/content/drive/MyDrive/2022/Urgenciastacotalpa2022.csv", delimiter=',', header=None, names=col_n:
datos1.head()
```

	ID	CLUES	fechaalta	EDAD	CVEEDAD	SEXO	ENTRESIDENCIA	MUNRESIDENCIA	DERHAB	TIPOURGENCIA	...	MP	AF
0	6408201	TCSSA004301	2022-03-04	11	4	1	27	15	0	2	...	8	
1	6408202	TCSSA004301	2022-03-08	45	5	2	27	15	0	2	...	8	
2	6408227	TCSSA004301	2022-12-03	21	5	2	27	15	0	2	...	8	
3	6408260	TCSSA004301	2022-03-	26	5	2	27	15	0	2	...	8	

*Nota.* Elaboración propia.

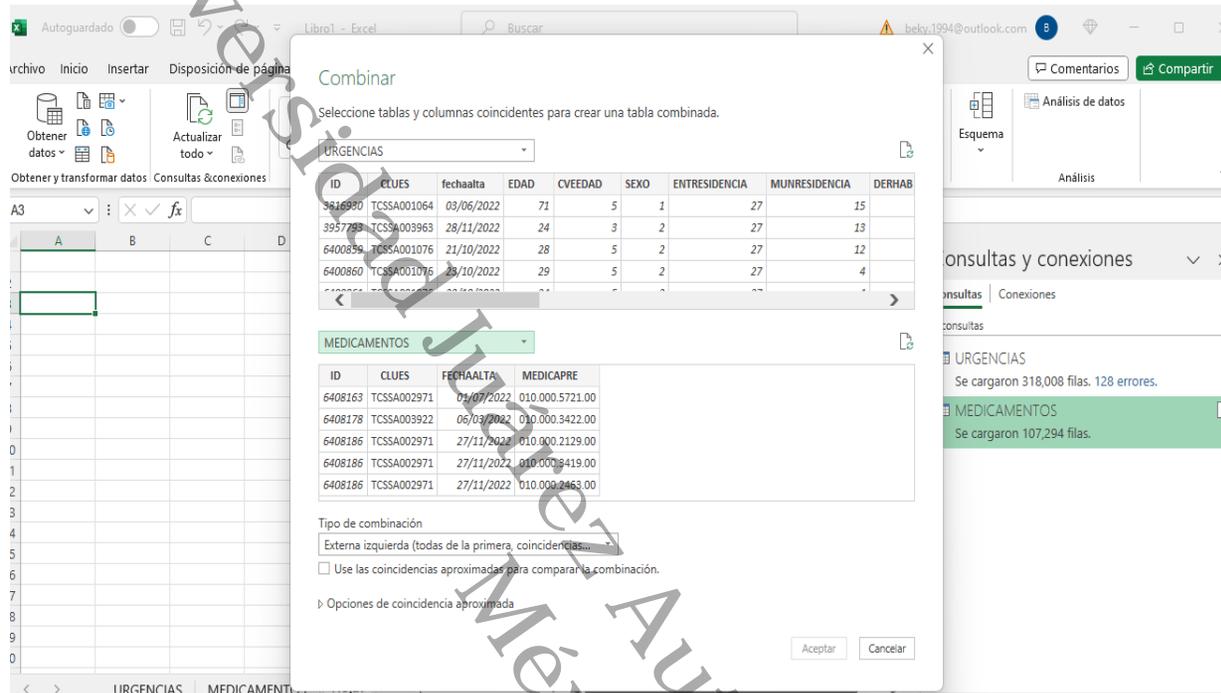
Se crearon nueve archivos con la extensión *csv*, correspondientes a las tablas de URGENCIAS, AFECCIONES Y MEDICAMENTOS de los años 2022, 2021 y 200.

El segundo paso que se realizó para la obtención del *dataset* fue la combinación de cada uno de los archivos *csv*, para realizar este proceso se utilizó Microsoft Excel como herramienta principal.

Para realizar este procedimiento se utilizó la Clave ID de cada registro para evitar duplicidad o pérdida de los datos. En la figura 12 se observa parte del proceso que se utilizó para la combinación de las tablas.

Figura 12

### Procedimiento para la combinación de las tablas



Nota. Elaboración propia.

Después de realizar la combinación de las tablas se obtuvo un archivo con un total de 25 variables. Posteriormente se eliminaron las variables que no se utilizaron para este proyecto, el resultado que se obtuvo fue un *dataset* constituido por un total de siete variables las cuales se describen a continuación:

- EDAD: edad cumplida del paciente
- SEXO: sexo del paciente, puede ser masculino o femenino
- TIPOURGENCIA: Se refiere si el paciente fue atendido como urgencia calificada o no calificada
- AFECPRIN: la afección principal del paciente
- AFEC: afecciones de comorbilidad del paciente
- MEDICAMENTO: nombre del medicamento que se le brindó al paciente

- **DESCRIPCION:** Describe el tipo de medicamento que se le prescribió al paciente. Puede ser inyectable, tabletas jarabe, etc.

En la figura 13 se muestran las variables que contiene el *dataset* que se diseñó para entrenar el modelo de minería de datos. El *dataset* contiene un total de 7,899 registros.

**Figura 13**

*Dataset utilizado en el proyecto*

EDAD	SEXO	TIPOURGENCIA	AFECPRIN	AFEC	MEDICAMENTOS	DESCRIPCION
17	1	2	DENGUE, NO ESPECIFICADO	DENGUE, NO ESPECIFICADO	PARACETAMOL	ENVASE CON CUATRO FRASCOS AMI
59	2	2	DIABETES MELLITUS NO ESPECIFI	OTRAS GASTROENTERITIS Y COLITIS I	KETOROLACO	FRASCO AMPULA O AMPOLLETA - C/
59	2	2	DIABETES MELLITUS NO ESPECIFI	OTRAS GASTROENTERITIS Y COLITIS I	SOLUCION HARTMANN	ENVASE CON 500 ML. MILIEQUIVALI
65	2	2	DEPLECIÃ" N DEL VOLUMEN	OTRAS GASTROENTERITIS Y COLITIS I	METRONIDAZOL	ENVASE CON 100 ML. - CADA 100 MI
3	2	2	DEPLECIÃ" N DEL VOLUMEN	OTRAS GASTROENTERITIS Y COLITIS I	BUTILHIOSCINA O HIOSCI	AMPOLLETA - CADA AMPOLLETA CO
65	2	2	DEPLECIÃ" N DEL VOLUMEN	OTRAS GASTROENTERITIS Y COLITIS I	METOCLOPRAMIDA	AMPOLLETA - CADA AMPOLLETA CO
3	2	2	DEPLECIÃ" N DEL VOLUMEN	OTRAS GASTROENTERITIS Y COLITIS I	METRONIDAZOL	ENVASE CON 2 AMPOLLETAS O FRAS
65	2	2	DEPLECIÃ" N DEL VOLUMEN	OTRAS GASTROENTERITIS Y COLITIS I	KETOROLACO	FRASCO AMPULA O AMPOLLETA - C/
3	2	2	DEPLECIÃ" N DEL VOLUMEN	OTRAS GASTROENTERITIS Y COLITIS I	SOLUCION HARTMANN	ENVASE CON 1000 ML. MILIEQUIVAI
19	1	2	FARINGITIS AGUDA, NO ESPECIFIC	OTRAS GASTROENTERITIS Y COLITIS I	LORATADINA	ENVASE CON 20 TABLETAS O GRAGE
19	1	2	FARINGITIS AGUDA, NO ESPECIFIC	OTRAS GASTROENTERITIS Y COLITIS I	OMEPRAZOL	ENVASE CON 7 TABLETAS O GRAGEA
82	2	2	ENFERMEDAD PULMONAR OBSTR	OTRAS GASTROENTERITIS Y COLITIS I	METRONIDAZOL	MILIGRAMO - CADA 100 ML CONTIE
28	2	2	OTROS DOLORES ABDOMINALES	OTRAS GASTROENTERITIS Y COLITIS I	METAMIZOL SODICO	ENVASE CON 3 AMPOLLETAS CON 2
28	2	2	OTROS DOLORES ABDOMINALES	OTRAS GASTROENTERITIS Y COLITIS I	BUTILHIOSCINA O HIOSCI	ENVASE CON 3 AMPOLLETAS DE 1 M
20	?	1	SUPERVISIÃ" N DE EMBARAZO NO	OTRAS GASTROENTERITIS Y COLITIS I	PARACETAMOL	ENVASE CON 10 TABLETAS - CADA T

*Nota.* Elaboración propia.

### 3.3 Fase 3: Evaluación de tres técnicas de minería de datos

Esta fase fue la más importante dentro de la metodología ya que en este paso se llevó a cabo el análisis del *dataset* con las técnicas de agrupamiento, arboles de decisión y reglas de asociación. Después de realizar el análisis de los datos con cada una de estas técnicas, se determinó la técnica que se utilizaría para la creación del modelo de minería de datos.

Para realizar el análisis se utilizó *Google Colab*, que es una herramienta gratuita para ejecutar código *Python* en la nube. A continuación, se describen los resultados de los análisis con cada una de las técnicas.

#### 3.3.1 Técnicas de agrupamiento

La técnica de agrupamiento o también conocida como *clustering* en la minería de datos, es un método de aprendizaje no supervisado que tiene como objetivo organizar los datos

analizados en grupos que tienen elemento o características similares entre sí. Con esta técnica se utilizó el algoritmo *K-means*.

A continuación, se muestran los resultados obtenidos con esta técnica:

El primer paso en la implementación del algoritmo *k-means* es la obtención de la gráfica de codo donde se puede obtener el número de *cluster* óptimos para realizar la agrupación de los datos. En la figura 14 se muestra el código que muestra los *cluster* obtenidos mediante la gráfica de codos.

### Figura 14

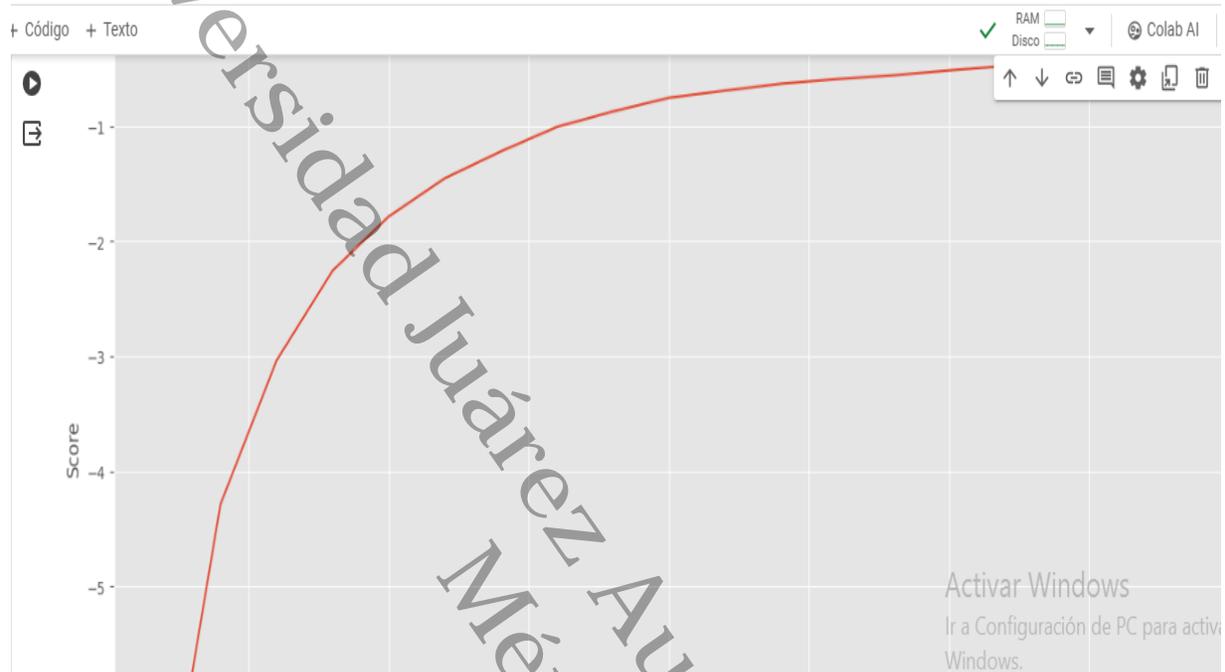
*Código implementado para obtener la gráfica de codos*

```
▶ Nc = range(1, 20)
kmeans = [KMeans(n_clusters=i) for i in Nc]
kmeans
score = [kmeans[i].fit(X).score(X) for i in range(len(kmeans))]
score
plt.plot(Nc, score)
plt.xlabel('Number of Clusters')
plt.ylabel('Score')
plt.title('Elbow Curve')
plt.show()
```

*Nota.* Elaboración propia

La figura 15 muestra la gráfica de codo, donde se observa que el resultado obtenido fueron 5 *clusters*.

**Figura 15**  
Gráfica de codo



*Nota.* Elaboración propia

En la figura 15 que representa la gráfica de codo podemos observar que la línea muestra un quiebre en el valor 5, lo cual nos indica que el número de *clusters* es cinco para la segmentación mediante *K-means*.

Después de identificar el número de *clusters*, se implementó el código para localizar los *centroides* de cada *clusters*. En la figura 16 se muestra el código ejecutado para localizar *centroides* de cada *clusters*.

**Figura 16**

*Código implementado para la obtención de los centroides*

```
kmeans = KMeans(n_clusters=5).fit(X)
centroids = kmeans.cluster_centers_
print(centroids)
```

*Nota.* Elaboración propia

La tabla 3 muestra los *centroides* que fueron localizados.

**Tabla 3**

*Centroides localizados*

<b>Centroides</b>				<b>Cluster</b>
90.38362069	510.91954023	44.49281609	0.62284483	Cluster1
403.50657895	79.49342105	37.97944079	0.63075658	Cluster2
97.83870968	80.2748139	41.07320099	0.64143921	Cluster3
396.2379936	437.39967983	36.80096051	0.54002134	Cluster4
202.31430155	257.66851441	36.44733925	0.74445676	Cluster5

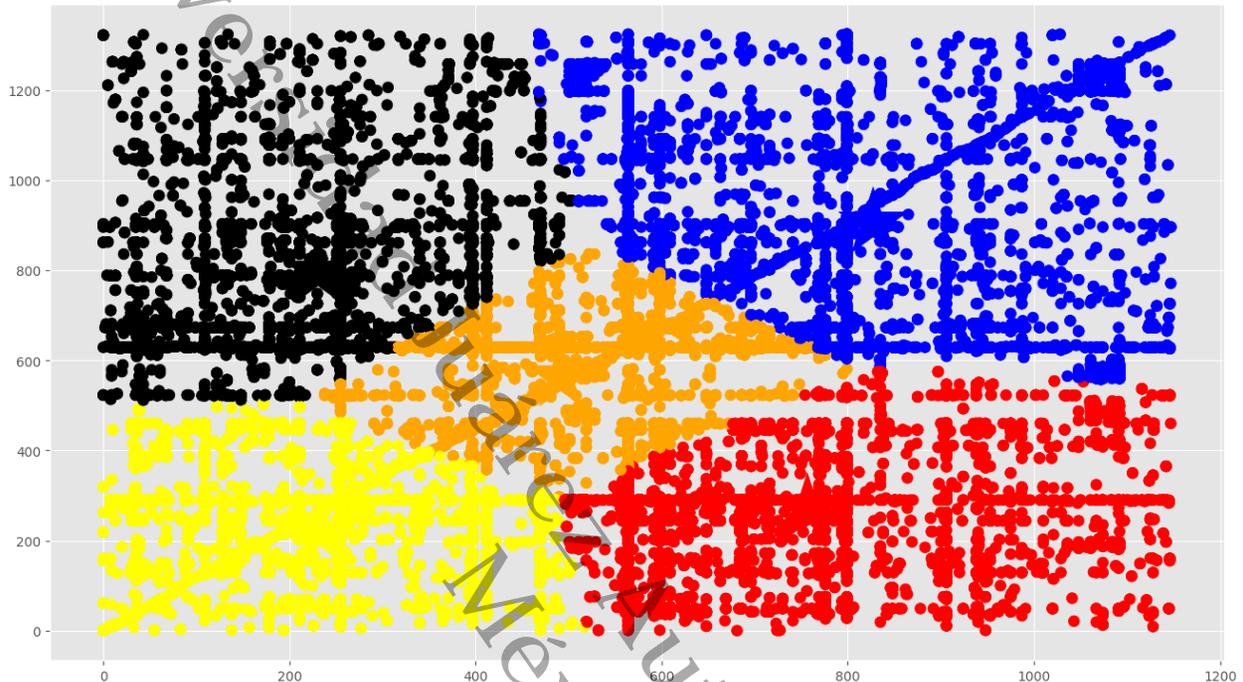
*Nota:* elaboración propia

La primera columna contiene los valores de la característica "Afecprin", la segunda columna contiene los valores de la característica "Afec", la tercera columna contiene los valores de la característica "Edad" y la cuarta columna contiene los valores de la característica "Sexo".

En la figura 17 se muestra la gráfica de la distribución de los *centroides* y la distribución de cada uno de los *clusters*.

**Figura 17**

Gráfica de la distribución de los cluster y centroides identificados



Nota. Elaboración propia.

Después de realizar el análisis de los *clusters* localizados se encontraron características que pueden ser útiles para determinar un tratamiento médico, principalmente considerando la diversidad de afecciones y la cantidad de medicamentos que fueron analizados. En conclusión, con la aplicación de esta técnica, se determinó que los datos pueden ser útiles para crear el modelo que determine tratamientos médicos.

### 3.3.2 Técnicas de reglas de asociación

Para la evaluación de la técnica de reglas de asociación se utilizó el algoritmo *Apriori*, este es uno de los algoritmos más utilizados en esta técnica, se utiliza para encontrar asociaciones existentes en un gran volumen de datos. Su principal función es identificar reglas de asociación, como aquellas que determinan la probabilidad de que ciertos ítems aparezcan juntos en una transacción.

Al realizar el análisis del *dataset* con la técnica de reglas de asociación se encontraron los siguientes resultados:

Primero se identificó la distribución del número de ítems por transacción, en la figura 19 se muestra el código escrito para realizar este procedimiento.

### Figura 18

*Distribución del número de ítems por transacción*

```
# Distribución del número de ítems por transacción
# -----
display(dataframe.groupby('AFEC')['MEDICAMENTOS'].size().describe(percentiles=[.25, .5, .75, .9]))

fig, ax = plt.subplots(figsize=(7, 3))
dataframe.groupby('AFECPRIN')['MEDICAMENTOS'].size().plot.hist(ax=ax)
ax.set_title('Distribución del tamaño de las transacciones');
ax.set_xlabel('Número de ítems');
```

*Nota.* Elaboración propia.

Los resultados que se obtuvieron al ejecutar el código se muestran en la figura 20.

### Figura 19

*Número de transacciones*

MEDICAMENTOS	
<b>count</b>	556.000000
<b>mean</b>	14.221223
<b>std</b>	51.930685
<b>min</b>	1.000000
<b>25%</b>	2.000000
<b>50%</b>	3.000000
<b>75%</b>	7.000000
<b>90%</b>	19.000000
<b>max</b>	703.000000

*Nota.* Elaboración propia.

Al analizar estos resultados se identificaron 556 transacciones.

**Mean (Media):** 14.22: en promedio, hay aproximadamente 14 medicamentos asociados en cada transacción o conjunto identificado como frecuente. Esto indica que, en general, las reglas de asociación involucran una combinación de varios medicamentos.

**Standard Deviation** (Desviación estándar): 51.93: la desviación estándar es relativamente alta, lo que sugiere una gran variabilidad en el número de medicamentos involucrados en las reglas de asociación. Esta dispersión indica que algunas transacciones contienen muchas más combinaciones de medicamentos que otras.

**Min (Mínimo):** 1.00: la transacción más pequeña (conjunto de medicamentos) involucra solo un medicamento. Esto puede representar una regla de asociación mínima o un medicamento que aparece frecuentemente solo.

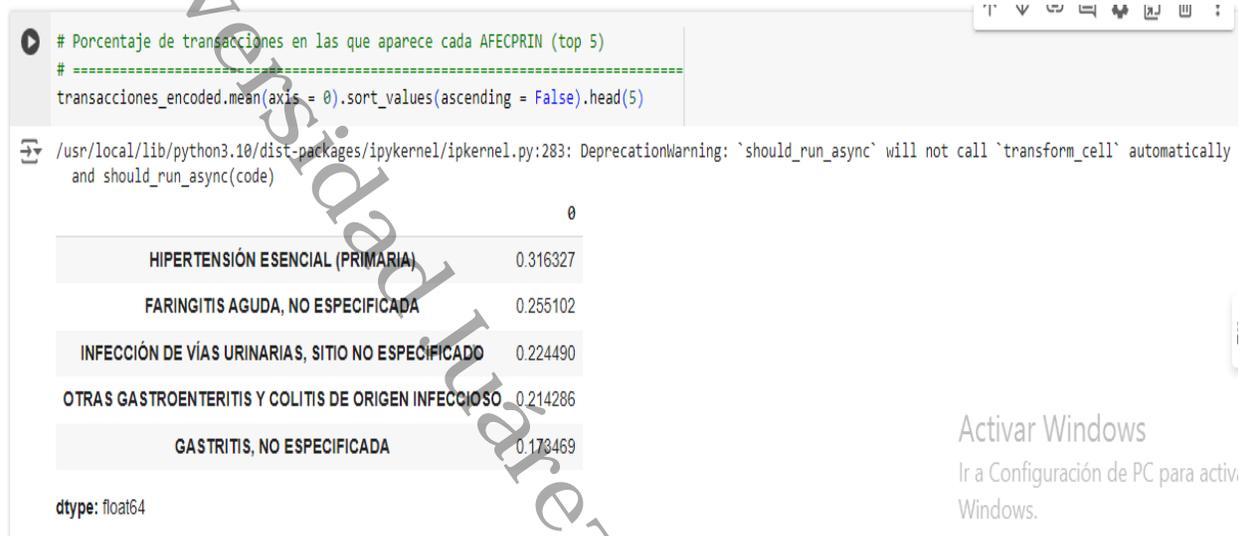
**Percentiles (25%, 50%, 75%, 90%):**

- El 25% de las transacciones contienen 2 o menos medicamentos.
- 50% la mitad de las transacciones contienen 3 o menos medicamentos. Esto indica que, aunque hay una gran variabilidad en el número de medicamentos, la mayoría de las transacciones involucran combinaciones relativamente pequeñas.
- El 75% de las transacciones contienen 7 o menos medicamentos. Esto confirma que la mayoría de las combinaciones no son extremadamente grandes.
- El 90% de las transacciones tienen hasta 19 medicamentos, indicando que solo una pequeña fracción de las combinaciones involucran un número significativamente alto de medicamentos.

**Max (Máximo):** la transacción más grande identificada por el algoritmo *Apriori* contiene 703 medicamentos. Este valor extremadamente alto podría ser un caso atípico o representar una regla de asociación que incluye una gran variedad de medicamentos. Se identificó la frecuencia relativa o la proporción de distintas afecciones en el conjunto de datos. En la figura 21 se observa cuáles son las cinco principales afecciones del *dataset*.

**Figura 20**

*Top 5 de Afecciones principales*



*Nota.* Elaboración propia.

Esta distribución ayuda a obtener información valiosa sobre la popularidad de los ítems que conforman el conjunto de datos y cómo se combinan para obtener las reglas de asociación.

El siguiente paso que se realizó en la creación de este modelo fue la determinación de las reglas de asociación. Después de escribir y ejecutar el código para la detección de las reglas de asociación se generaron un total 573218 reglas, con una confianza mínima de 0.7 y una confianza máxima de 1.0. En la figura 22 se muestra el código escrito para localizar las reglas de asociación en este modelo.

**Figura 21**

Código escrito para identificar las reglas de asociación generadas

```
# Identificación de itemsets frecuentes
# =====
soporte = 30 / transacciones_encoded.shape[0]
itemsets_frecuentes = apriori(transacciones_encoded, min_support=soporte, use_colnames=True)

# Crear reglas de asociación (confianza mínima del 70% para que una regla sea seleccionada)
# =====
confianza = 0.7
reglas = association_rules(itemsets_frecuentes, metric="confidence", min_threshold=confianza)

print(f"Número de reglas generadas: {len(reglas)}")
print(f"Confianza mínima: {reglas['confidence'].min()}")
print(f"Confianza máxima: {reglas['confidence'].max()}")
reglas.sort_values(by='confidence').head(5)
```

*Nota.* Elaboración propia.

Al realizar el análisis de los datos con esta técnica se encontraron reglas claras para la clasificación y prescripción de medicamentos, y esto se refleja en con el porcentaje que se obtuvo de confianza mínima (70%).

Para lograr un resultado óptimo es necesario aplicar métricas adicionales como el "*lift*" (elevación) o el "*support*" (soporte) para priorizar las reglas más útiles. También se podría considerar un análisis de las reglas más relevantes o enfocarse en aquellas que mejoran la eficacia clínica, esto debido al gran número de reglas de asociación que se generaron.

### 3.3.3 Técnicas de árboles de decisión

La última técnica que se evaluó fue la técnica de árboles de decisión. Esta técnica es un método en la minería de datos y el aprendizaje automático que se utiliza para la toma de decisiones y realizar predicciones. La técnica de árbol de decisión tiene una estructura en forma de árbol y se divide en nodos, ramas y divisiones. Al realizar el análisis del *dataset* con la técnica de árboles de decisión se encontraron los siguientes resultados:

Uno de los primeros pasos que se realizó con esta técnica fue dividir el conjunto de datos en características ( $x$ ) y la variable objetivo ( $y$ ). En la figura 23 se muestra el código escrito para realizar este procedimiento.

**Figura 22**

Código para dividir el conjunto de datos en características y variable objetivo

```
# Definir las columnas de características sin la variable objetivo
feature_cols = ['Edad', 'Sexo', 'Tipourgencia', 'Afecprin', 'Afec', 'Descripcion']

# Dividir el conjunto de datos en características (X) y la variable objetivo (y)
X = dataframe[feature_cols] # Características
y = dataframe['Medicamento'] # Variable objetivo
```

Nota. Elaboración propia.

Se realizó la división del conjunto de datos en un conjunto de entrenamiento asignando un 70% de los datos para el entrenamiento del modelo y el 30% para la prueba, se utilizó la función `train_test_split` de `sklearn.model_selection`. En la figura 24 se puede observar el código que se escribió para realizarlo.

**Figura 23**

Código para dividir el conjunto de datos en características y etiquetas de entrenamiento

```
# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1) # 70% training and 30% test
```

Nota. Elaboración propia.

`X_train`, `X_test`: representan las características del conjunto de entrenamiento y de prueba, respectivamente.

`y_train`, `y_test`: Son las etiquetas o la variable objetivo del conjunto de entrenamiento y de prueba, respectivamente.

A continuación, se escribió el código para entrenar el clasificador de árbol de decisión, para realizar este procedimiento se utilizó `DecisionTreeClassifier` de la biblioteca `sklearn`, y luego se evalúa el modelo utilizando varias métricas como la matriz de confusión, el informe de clasificación y la precisión del modelo. Finalmente, visualiza

el árbol de decisión. En la figura 25 se observa el código escrito para realizar este procedimiento.

### Figura 24

*Código utilizado para entrenar el clasificador del árbol de decisión*

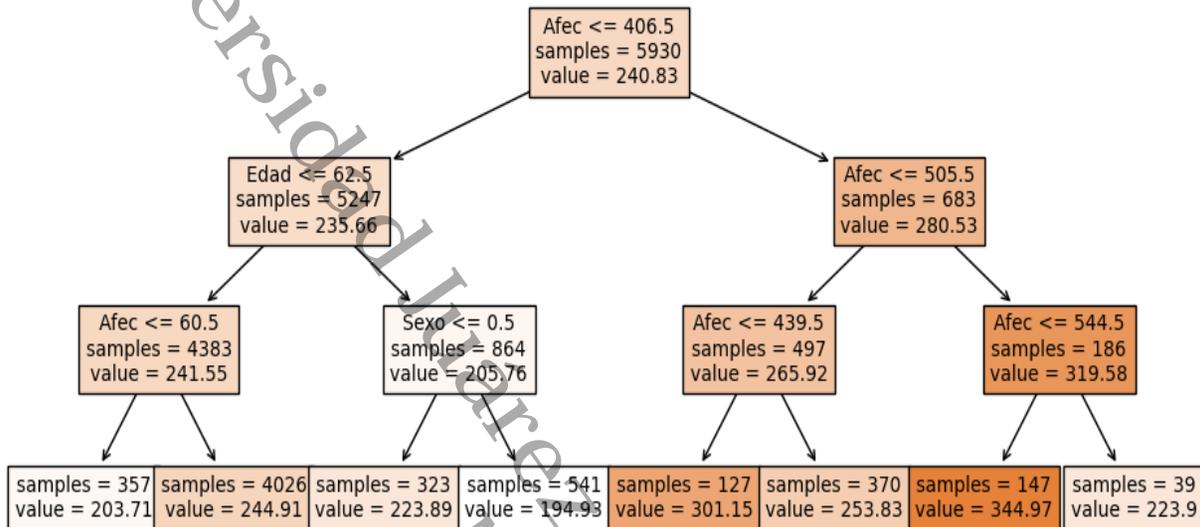
```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.tree import plot_tree
import matplotlib.pyplot as plt
# Crear el clasificador de árbol de decisión
clf = DecisionTreeClassifier(random_state=1)
# Entrenar el clasificador
clf.fit(X_train, y_train)
# Predecir la respuesta para el conjunto de prueba
y_pred = clf.predict(X_test)
# Evaluación del modelo
print("Matriz de confusión:")
print(confusion_matrix(y_test, y_pred))
print("\nInforme de clasificación:")
print(classification_report(y_test, y_pred))
print("\nPrecisión del modelo:")
print(accuracy_score(y_test, y_pred))
# Visualización del árbol de decisión
plt.figure(figsize=(20,10))
plot_tree(clf, filled=True, feature_names=X_train.columns, class_names=True)
plt.show()
```

*Nota.* Elaboración propia

Después de ejecutar este código se obtuvo una precisión del modelo de 0.967551622418879. Es bastante alta, lo que indica que el clasificador de árbol de decisión realiza predicciones correctas en casi el 97% de los casos. La estructura del árbol que se obtuvo se muestra en la figura 26.

**Figura 25**

*Estructura del árbol*



*Nota.* Elaboración propia.

Las variables Afec, Edad y Sexo son las características de más relevancia en el modelo para predecir la variable objetivo. En la figura 26 se observa el porcentaje de importancia de cada una de estas variables. Cuanto mayor sea el valor de importancia, más influencia tiene ese predictor en las decisiones del modelo.

**Figura 26**

*Cuadro de predictores e importancia del modelo*

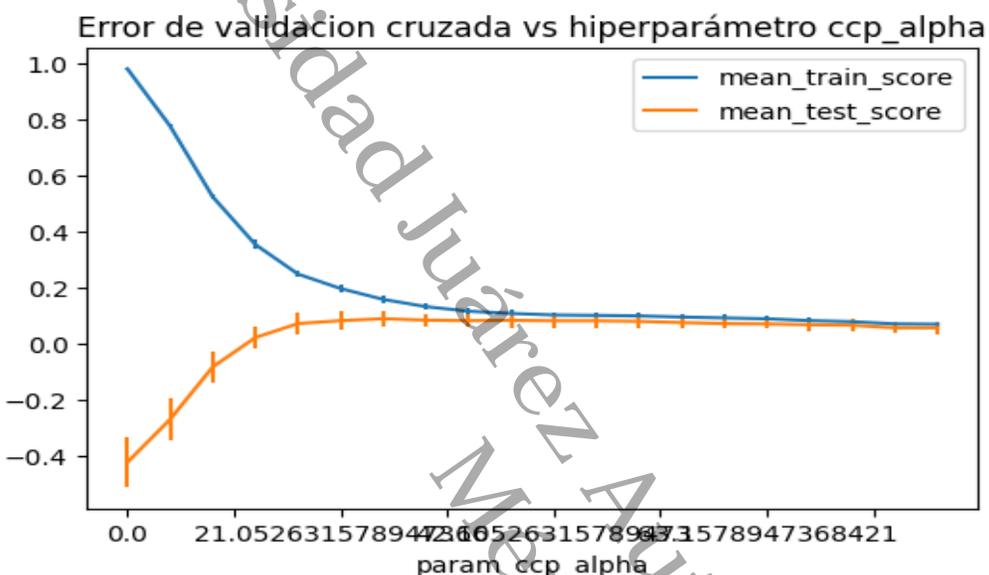
	predictor	importancia
3	Afec	0.720866
0	Edad	0.235852
1	Sexo	0.043282
2	Tipourgencia	0.000000
4	Medicamento	0.000000
5	Descripcion	0.000000

*Nota.* Elaboración propia.

Se realizó la poda del árbol obtenido utilizando validación cruzada, en la figura 28 se muestran el resultado obtenido de este procedimiento.

### Figura 27

Cuadro de predictores y la importancia del modelo



Nota. Elaboración propia.

La curva de *mean\_train\_score*: muestra el rendimiento promedio del modelo en el conjunto de entrenamiento para cada valor de *ccp\_alpha*.

La curva de *mean\_test\_score*: muestra el rendimiento promedio del modelo en el conjunto de prueba durante la validación cruzada. El punto donde esta curva alcanza su máximo es el *ccp\_alpha* óptimo, el cual minimiza el error de generalización.

Con el proceso de poda que se aplicó al modelo se buscó minimizar errores de generalización para asegurarse que el modelo no solo funcione bien con los datos de entrenamiento, sino también con nuevos datos. Al realizar este procedimiento se observó que el modelo no está generalizando bien los datos de prueba. El árbol de decisión entrenado es un buen punto de partida para analizar decisiones médicas, especialmente debido a su transparencia e interpretabilidad. Sin embargo, para determinar tratamientos

médicos específicos de manera confiable, se necesitaría una validación adicional, posiblemente una mayor complejidad del modelo y datos más detallados.

Universidad Juárez Autónoma de Tabasco.  
México.

## Capítulo 4. Resultados

Al realizar la evaluación de las tres técnicas de minería de datos se determinó que la técnica de agrupamiento y el algoritmo *k-means* ofrecía una mejor alternativa para desarrollar el modelo de minería de datos para implementarlo en un sistema de recomendaciones médicas, esto de acuerdo a los resultados que se obtuvieron del análisis de los datos con esta técnica.

Se consideraron cinco métricas para comparar el rendimiento de las técnicas de árboles de decisión, reglas de asociación y agrupamiento en la predicción de medicamentos. En la tabla 4 se muestra la comparación de los resultados obtenidos con cada una de las técnicas.

**Tabla 4**

*Resultados de la evaluación de las tres técnicas de minería de datos*

Técnica	Precisión (%)	Cobertura (%)	F1 Score	Tiempo de procesamiento (s)	Casos correctamente diagnosticados (%)
Árboles de decisión	75.6	72.3	0.4	12.5	70.1
Reglas de asociación	68.2	78.5	0.73	18.0	65.4
Agrupamiento	82.3	85.7	0.84	8.3	79.8

*Nota.* Elaboración propia.

- **Precisión:** mide qué tan bien el modelo predice los medicamentos correctos en función de los síntomas y las afecciones del paciente.
- **Cobertura:** indica qué porcentaje de las recomendaciones correctas posibles fue capturado por la técnica.
- **F1 Score:** proporciona una métrica equilibrada que combina precisión y cobertura.
- **Tiempo de procesamiento:** evalúa la eficiencia de cada técnica en términos de rapidez.

- **Casos correctamente diagnosticados:** muestra el porcentaje de casos en los que el modelo predijo correctamente el medicamento para los pacientes.

En este apartado se describe el código que se escribió para crear el modelo, la interfaz gráfica que tiene el sistema de recomendaciones y el resultado de las evaluaciones que se realizaron al sistema.

#### **4.1 Código escrito para desarrollar el modelo**

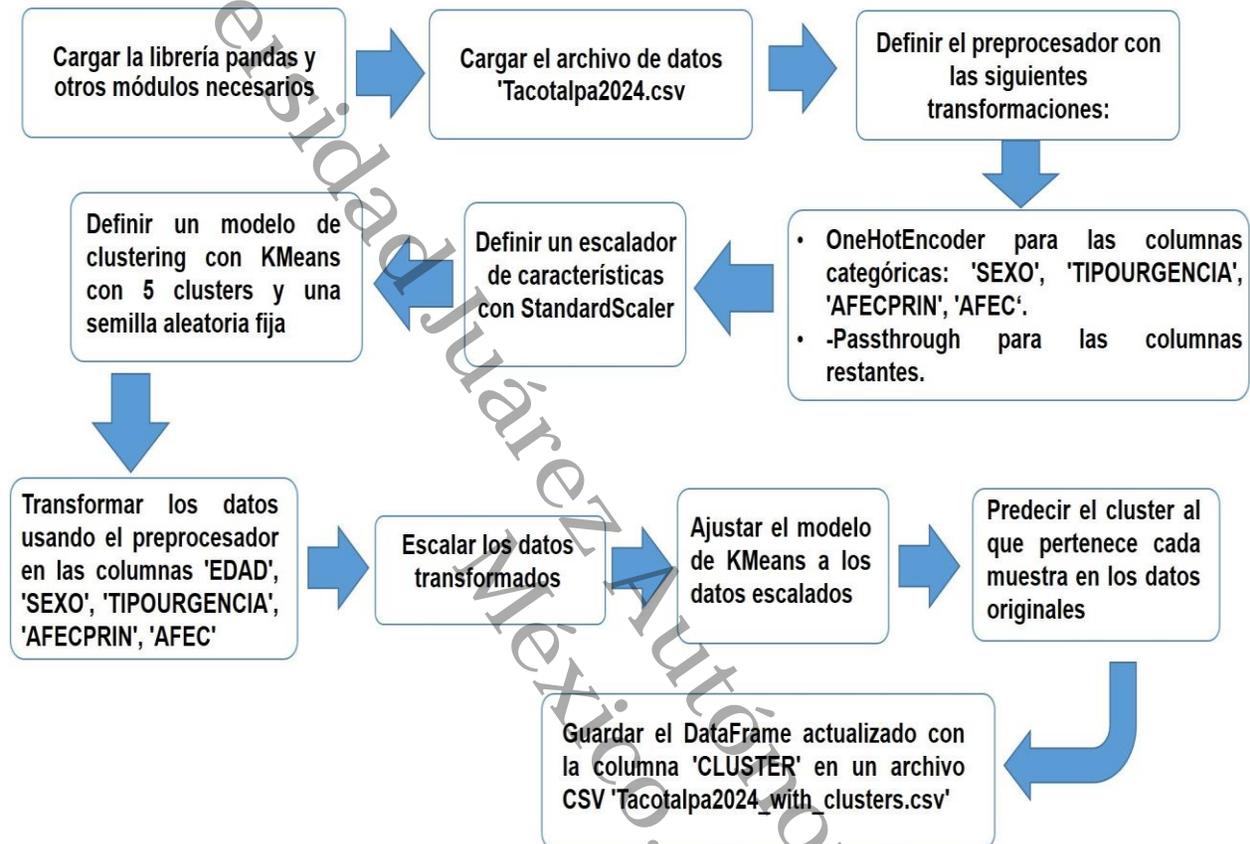
Para generar el código del modelo de minería de datos se utilizó la herramienta de *Visual Studio Code*, el código se dividió en cuatro partes, la primera parte es un script para cargar y preprocesar el conjunto de datos, en la segunda parte del código se definen las funciones para cargar datos y modelos, en la tercera parte del código se creó la interfaz de usuario usando *Streamlit* y en la cuarta parte se escribió el código que predice los medicamentos y muestra los resultados en una tabla, a continuación, se muestra el pseudocódigo que describe cada una de las partes del código principal.

##### **4.1.1 Cargar y preprocesar los datos (primera parte)**

En la figura 28 se muestra cada uno de los procesos realizados en la primera parte del código.

**Figura 28**

Seudocódigo que muestra la primera parte del código escrito para este proyecto



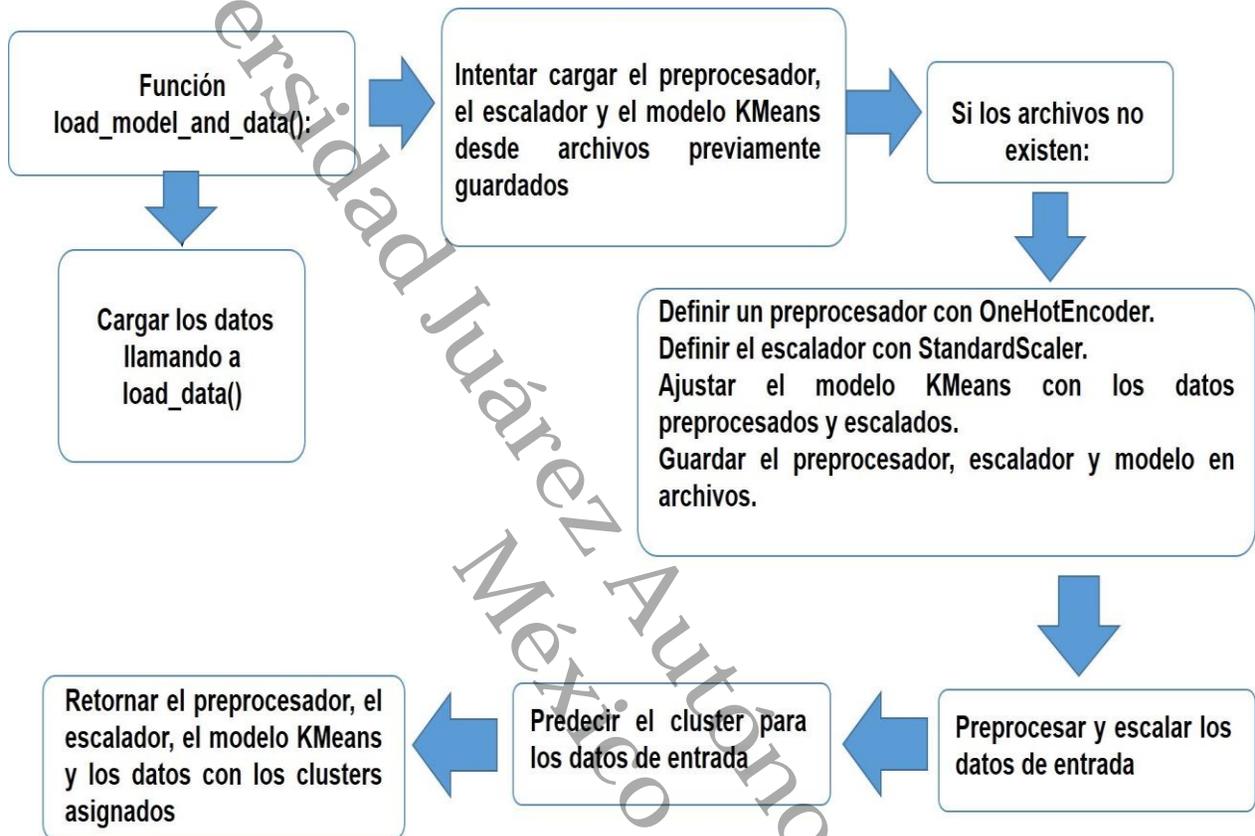
Nota. Elaboración propia.

#### 4.1.2 Definir funciones para cargar datos y modelos (segunda parte)

En la figura 29 se muestra cada uno de los procesos realizados en la segunda parte del código.

**Figura 29**

Seudocódigo que muestra la segunda parte del código escrito para este proyecto



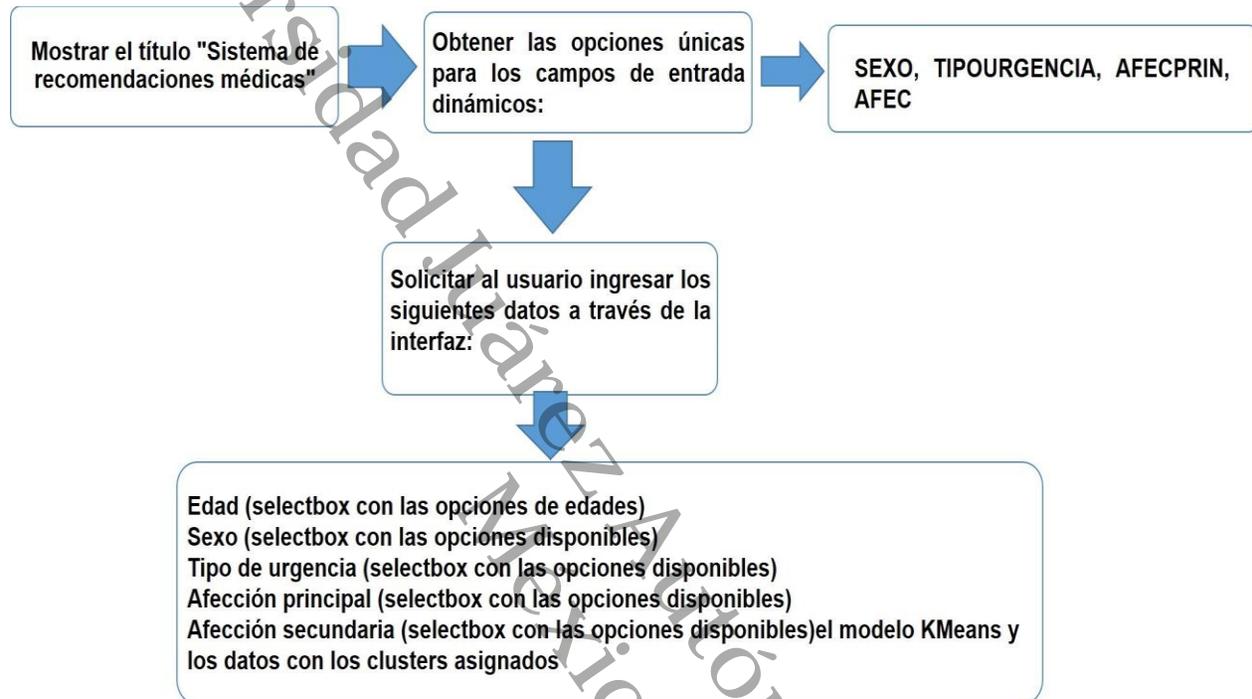
Nota. Elaboración propia.

#### 4.1.3 Configurar la interfaz de usuario con *Streamlit*

En la figura 30 se muestra cada uno de los procesos realizados en la tercera parte del código.

**Figura 30**

*Seudocódigo que muestra la tercera parte del código escrito para este proyecto*



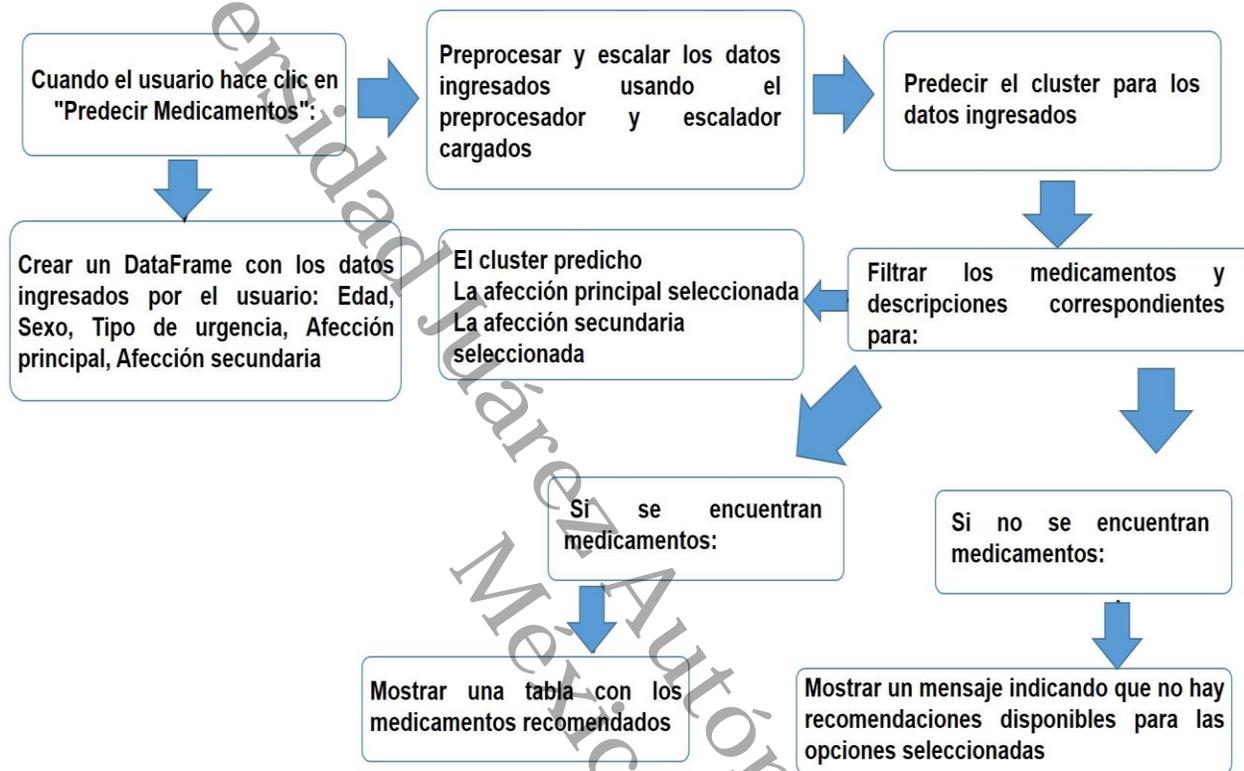
*Nota.* Elaboración propia

#### **4.1.4 Predecir medicamentos y mostrar resultados**

En la figura 31 se muestra cada uno de los procesos realizados en la cuarta parte del código.

**Figura 31**

Seudocódigo que muestra la cuarta parte del código escrito para este proyecto



*Nota.* Elaboración propia

## 4.2 Interfaz gráfica del sistema de recomendaciones médicas

La interfaz gráfica de la aplicación esta constituida por cinco campo de entradas que tienen que ser seleccionadas en un selectbox, a continuación se describen cada una de las entradas:

**Edad:** este campo permite al usuario seleccionar la edad del paciente. Puede tener un rango de 0 a 100 años.

Tipo de campo: desplegable con valores numéricos.

Función: la edad del paciente es una variable importante para definir el perfil del paciente y puede influir en las recomendaciones de medicamentos, ya que ciertos tratamientos varían según la edad.

**Sexo:** este campo permite seleccionar el sexo del paciente, 1 para Masculino y 2 para Femenino.

Tipo de campo: desplegable con opciones numéricas predefinidas.

Función: el sexo del paciente es otra variable importante ya que algunos medicamentos y tratamientos pueden estar diferenciados por género.

**Tipo de Urgencia:** este campo permite al usuario seleccionar el tipo de urgencia médica con la que llega el paciente, esta puede ser 1 para Urgencia Calificada y 2 para Urgencia No Calificada.

Tipo de campo: desplegable con opciones numéricas.

Función: el tipo de urgencia afecta el tratamiento y las recomendaciones médicas, ya que algunas afecciones pueden requerir tratamiento inmediato mientras que otras no.

**Afección Principal:** este campo permite seleccionar la afección o condición médica principal que presenta el paciente. La afección es seleccionada mediante un selectbox donde están registradas todas afecciones contenidas en el dataset.

Tipo de campo: desplegable con opciones categóricas.

Función: la afección principal es la enfermedad o condición médica más importante a la que se debe prestar atención en el diagnóstico y tratamiento. Esto influirá directamente en la predicción de los medicamentos.

**Afección Secundaria:** este campo permite seleccionar una afección o condición médica secundaria que también puede estar presente en el paciente. La afección secundaria es seleccionada mediante un selectbox donde están registradas todas afecciones contenidas en el dataset.

Tipo de campo: desplegable con opciones categóricas.

Función: la afección secundaria también influye en las recomendaciones de tratamiento, ya que en algunos casos se requiere ajustar la medicación o el tratamiento en función de la coexistencia de varias condiciones.

**Botón "Predecir Medicamentos":** este es el botón de acción que permite ejecutar el proceso de predicción de medicamentos con base en los datos ingresados.

Función: una vez que el usuario ha seleccionado todas las opciones en los campos anteriores, presionar este botón ejecuta el modelo de *K-means* para predecir el cluster al que pertenece el paciente y, en función de ello, se muestran las recomendaciones de medicamentos.

En la figura 32 se observan cada uno de los campos descritos anteriormente.

### Figura 32

*Interfaz gráfica del sistema de recomendaciones médicas*

## Sistema de recomendaciones médicas

Edad

0

Sexo (1=Masculino, 2=Femenino)

1

Tipo de Urgencia (1=Urgencia Calificada, 2=Urgencia no Calificada)

2

Afección Principal

DENGUE, NO ESPECIFICADO

Afección Secundaria

DENGUE, NO ESPECIFICADO

Predecir Medicamentos

*Nota.* Elaboración propia.

## **4.3 Evaluaciones del sistema de recomendaciones médicas**

### **4.3.1 Evaluación de funcionalidad del sistema.**

Como parte de la validación del sistema de recomendaciones médicas, se realizó la evaluación por parte un médico general con experiencia en atención en el área de primer y segundo nivel, esta actividad se hizo con la finalidad de asegurarse que el sistema cumpliera con los requisitos clínicos y operativos en un entorno real.

La evaluación consistió en una serie de pruebas prácticas, en las cuales el médico ingreso datos de casos clínicos reales que él ha atendido para determinar si los medicamentos que el sistema le mostro coinciden con los que el como médico prescribiría a cada uno de los casos.

La importancia de esta evaluación radica en la necesidad de garantizar que el sistema no solo funcione correctamente desde un punto de vista técnico, sino que también sea útil y confiable para quienes estarán a cargo de su uso en la práctica clínica. Las observaciones y comentarios obtenidos fueron esenciales para ajustar las funcionalidades del sistema, mejorar su interfaz y asegurar que los resultados entregados coincidieran con las expectativas médicas y los protocolos clínicos vigentes.

A continuación, se mencionan cada una de las pruebas y observaciones que realizó el médico después de evaluar el sistema:

Prueba 1. Niño con faringitis aguda, El sistema recomendó los medicamentos que se deben utilizar para tratar esta afección.

Prueba 2. Enfermedad obstructiva crónica, de acuerdo con la edad y el tipo de urgencia el sistema mostró como resultados los medicamentos que se deben utilizar para tratar este padecimiento.

Prueba 3. Asma en los niños, el médico mencionó que el sistema únicamente mostro los medicamentos que son de rescate, ya que en la actualidad y de acuerdo con las guías GPC (Guías de Práctica Clínica) se deben prescribir otros medicamentos adicionales a los de rescate, el médico indicó que probablemente se deba a la fuente de datos que se utilizó para este proyecto.

Prueba 4. Diagnóstico de vaginitis aguda no mostro ningún medicamento.

Prueba 5. Cólico vesicular, el sistema dio como resultados los medicamentos que se utilizan para tratar este padecimiento.

#### **4.3.2 Comentarios finales del médico.**

“Yo considero que como el cuadro básico que se maneja en los hospitales regionales es limitado van a salir esos medicamentos, y están bien, se basan mucho en el cuadro básico del primer y segundo nivel de atención médica de la secretaria de Salud. El sistema muestra como resultados buenos medicamentos”.

Como resultado de esta evaluación, el médico consideró que el sistema cumplía adecuadamente con los objetivos propuestos, generando recomendaciones de medicamentos pertinentes para la mayoría de los casos. El resultado que el médico determinó en esta evaluación respalda la fiabilidad del sistema, lo que incrementa su potencial para ser implementado en situaciones reales de atención médica.

#### **4.3.3 Evaluación de la usabilidad del sistema.**

Después que el medico evaluara la funcionalidad del sistema, se le solicitó que valorara la usabilidad del sistema para que el médico realizara observaciones en cuanto al funcionamiento de la interfaz gráfica del sistema, con el fin de detectar posibles errores en la operatividad del sistema y validar la precisión de las sugerencias terapéuticas.

. Las observaciones que el médico realizó en esta evaluación fueron determinantes para realizar las mejoras antes de la implementación final del sistema.

Comentarios del médico después de realizar la evaluación de usabilidad del sistema.

**Interfaz de usuario:** la interfaz es intuitiva y fácil de navegar. Los campos requeridos están bien organizados, lo que facilita su uso incluso para aquellos que no tienen experiencia previa con este tipo de sistemas.

**Proceso de introducción de datos:** la selección de opciones como la edad del paciente, sexo, tipo de urgencia y afecciones principales/secundarias es clara. La

disposición de los menús desplegados facilita el ingreso de información de manera rápida y sin confusiones.

**Velocidad y rendimiento:** el tiempo de respuesta del sistema es óptimo, generando resultados en cuestión de segundos tras ingresar los datos del paciente. Esto es crucial en contextos de urgencia médica.

De acuerdo a las observaciones que realizó el médico se puede determinar que el sistema funciona adecuadamente y podría tener una alta usabilidad en el área médica. El formato que se utilizó para realizar esta evaluación se puede encontrar en los anexos de este trabajo de investigación.

#### **4.3.4 Otros resultados**

Como resultado del trabajo realizado en esta investigación, se elaboraron y publicaron dos artículos en revistas indexadas y arbitradas, lo que contribuye significativamente a la difusión de los hallazgos y la validación del sistema de recomendaciones médicas desarrollado. Estas publicaciones no solo reflejan la importancia y relevancia del proyecto en el campo de la salud, sino que también fortalecen la credibilidad científica de la investigación al haber pasado por rigurosos procesos de revisión por pares.

Estas publicaciones refuerzan la importancia de los resultados obtenidos, validando el sistema desde una perspectiva tanto técnica como clínica, y abren la puerta a futuras investigaciones en el campo de las recomendaciones automatizadas en la atención médica. Los artículos también actúan como una plataforma de difusión de los avances tecnológicos logrados, lo que facilita la colaboración interdisciplinaria y fomenta el interés en la implementación de soluciones tecnológicas innovadoras en la medicina.

Las cartas de aceptación de los artículos publicados se encuentran en la parte de anexos de esta investigación.

## Capítulo 5. Conclusiones, recomendaciones y trabajos futuros

### 5.1 Conclusiones

El sistema de recomendaciones médicas desarrollado demostró ser efectivo al proporcionar sugerencias de medicamentos adecuadas en función de los datos del paciente, el médico que evaluó el sistema confirmó su utilidad, destacando su capacidad para apoyar la toma de decisiones clínicas.

La interfaz de usuario fue calificada por el médico que evaluó el sistema como intuitiva y fácil de usar, de acuerdo con los resultados de la evaluación el médico indicó que la integración de listas desplegables para la selección de variables como sexo, edad, tipo de urgencia, y afecciones permite una entrada de datos rápida y sin complicaciones.

Aunque el sistema funcionó bien con los datos disponibles, se reconoce que su aplicabilidad y precisión pueden ampliarse significativamente al incorporar una mayor cantidad de datos médicos, lo que permitirá obtener recomendaciones más personalizadas y específicas.

De acuerdo con las preguntas de investigación para este proyecto se puede afirmar que las variables seleccionadas para generar el *dataset* y entrenar el modelo generaron resultados con un 90% de efectividad.

De las tres técnicas de minería de datos que se utilizaron en este proyecto, fue la técnica de agrupamiento y el algoritmo *K-means* ya que fue la técnica que brindó resultados en un 90% más acertados cuando el modelo fue implementado en el sistema de apoyo a la toma de decisiones.

La validación a través de la evaluación del médico fue un paso clave en el proceso de desarrollo. Sus comentarios permitieron ajustar y optimizar el sistema para garantizar que se ajuste a las necesidades prácticas de los profesionales de la salud.

## 5.2 Recomendaciones

**Agregar más datos al *dataset*:** es recomendable incrementar las variables del *dataset*, variables que ayuden a que el sistema sea más efectivo, ya que actualmente los resultados obtenidos fueron de un 90% de efectividad. Esto permitiría mejorar la precisión de las recomendaciones.

**Incorporación de Inteligencia Artificial:** en futuras versiones del sistema, se podría considerar la integración de algoritmos de inteligencia artificial más avanzados, como redes neuronales, para mejorar la precisión de las recomendaciones de medicamentos, basándose en patrones complejos de datos.

## 5.3 Trabajos futuros

**Incorporación de datos en tiempo real:** un área de investigación futura consiste en la incorporación de datos en tiempo real, como el historial clínico digital del paciente o indicadores vitales en tiempo real. Esto permitiría que el sistema ofrezca recomendaciones más actualizadas y precisas.

**Adaptación a nuevas especialidades médicas:** se propone expandir el alcance del sistema para abarcar otras especialidades médicas, como pediatría, geriatría o enfermedades crónicas, lo que ampliaría su aplicabilidad a diferentes contextos clínicos.

**Interoperabilidad con otros sistemas:** sería importante explorar la interoperabilidad del sistema con otros *softwares* médicos y bases de datos electrónicas de pacientes.

<b>Alojamiento de la Tesis en el Repositorio Institucional</b>	
<b>Título de la tesis:</b>	Minería de Datos aplicada en un Sistema de Apoyo a la Toma de Decisiones Médicas
<b>Autor de la Tesis:</b>	Ismael Zarate Hernandez
<b>ORCID:</b>	<a href="https://orcid.org/0009-0003-5208-316X">https://orcid.org/0009-0003-5208-316X</a>
<b>Resumen de la tesis</b>	<p>Las complicaciones causadas por los errores en la prescripción de tratamientos médicos son un problema que pueden ocurrir en cualquiera de los niveles de atención de salud. Para los pacientes, una mala prescripción puede ocasionar desde reacciones alérgicas y toxicidad hasta la ineficacia del tratamiento, lo que puede agravar su estado de salud e incluso poner en riesgo su vida. Para los médicos los errores en la prescripción pueden afectar su credibilidad y provocar consecuencias legales o disciplinarias, especialmente si se demuestra que hubo negligencia o falta de actualización en sus conocimientos médicos. Esta investigación se realizó con la finalidad de desarrollar un modelo de minería de datos para implementarlo en un sistema de apoyo a la toma de decisiones en el área médica. Este sistema fue diseñado con la finalidad de asistir a los médicos del segundo nivel de atención médica en la determinación de tratamientos clínicos adecuados para los pacientes, con el propósito de minimizar el margen de error en la prescripción médica.</p>
<b>Palabras claves de la tesis</b>	Minería de datos, algoritmo <i>k-means</i> , <i>Clustering</i> , Errores de diagnósticos.
<b>Referencias citadas:</b>	En la siguiente página se muestran las referencias

## Referencias

- Acosta-Sánchez, R., Rosete-Suárez, A., & Rodríguez-Díaz, A. (2009). Predicción de pacientes diabéticos. Preprocesado para minería de datos. *Revista Cubana de Informática Médica*, 9(1).
- Ahlemeyer-Stubbe, A., & Coleman, S. (2014). *A practical guide to data mining for business and industry*. John Wiley & Sons.
- Belginova, S., Uvaliyeva, I. y Rustamov, S. (2019). *The application of data mining methods for the process of diagnosing diseases*. *Journal of Theoretical and Applied Information Technology*, 97(7), 1980 – 1985. <http://www.jatit.org/volumes/Vol97No7/13Vol97No7.pdf>
- Bellazzi R, Zupan B. (2008) *Predictive data mining in clinical medicine: current issues and guidelines*. *Int J Med Inform.* 77(2) 81-97. doi: 10.1016/j.ijmedinf.2006.11.006.
- Beltran, D., Povea, D. (2010) Rapidminer, Universidad Nacional Colombia facultad de ciencias económicas unidad de informática y comunicaciones bogotá d.c. <http://fce.unal.edu.co/media/files/UIFCE/Economia/RapidMiner.pdf>
- Brennan TA, Leape LL, Laird NM, et al. Incidencia de eventos adversos y negligencia en pacientes hospitalizados: resultados del Estudio de práctica médica de Harvard I. *N Engl J Med* 1991; 324:370-6
- Brito, L. (2008). Niveles de atención en salud: Estrategias y desafíos. *Revista de Salud Pública*, 10(3), 230-245.
- Comisión Estatal de Conciliación y Arbitraje Médico. *Primer Informe de Actividades 2022*. México, 2023. <https://tabasco.gob.mx/sites/default/files/users/cecamettabasco/1ERINFORMEDEACTIVIDADES.pdf>
- Cámara de diputados del Honorable congreso de la unión. (2010). *Ley federal de protección de datos personales en posesión de los particulares*. Diario oficial de la federación. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

- Cámara de diputados del Honorable congreso de la unión. (1984). *Ley general de salud*. Diario oficial de la federación. [https://dof.gob.mx/nota\\_detalle.php?codigo=4652777&fecha=07/02/1984](https://dof.gob.mx/nota_detalle.php?codigo=4652777&fecha=07/02/1984)
- Davila-Hernandez, F., Sánchez-Corales, Y. (2012). *Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas*. Revista Cubana de Informática Médica, 12(2),2-13.
- Daza-Vergaray, A. (2016). *Data Mining Minería de Datos*. Macro EIRL.
- Fayyad, U., Piatetsky-Shapiro, G. y Padhraic, S. (1996). *The KDD Process for Extracting Useful Knowledge from Volumenes of Dath*. Communications of the ACM, 39(11), 29. <https://dl.acm.org/doi/pdf/10.1145/240455.240464>.
- Fernández-Cantón, S. (2016). *El error médico como causa de muerte*. Boletín CONAMED, Volumen (número de la revista), 26 – 31. [http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin6/error\\_medico.pdf](http://www.conamed.gob.mx/gobmx/boletin/pdf/boletin6/error_medico.pdf)
- Flores Guerrero, D. (2019). *Aplicación de minería de datos para el pronóstico de la evolución de la diabetes en México [Tesis de maestría, Tecnológico Nacional de México]*. Repositorio Institucional del TecNM. <https://rinacional.tecnm.mx/handle/TecNM/1377>
- Haro-Rivera, S. (2016) *Técnicas de clasificación en minería de datos y software Orange canvas. Aplicación con datos meteorológicos. [Máster en estadística aplicada]*. Universidad de Granada.
- Hernandez, Orallo, J., Ramirez, Quintana, J., Ferri, Ramirez, C. (2005) *Introducción a la Minería de Datos*. Departamento de Sistemas Informáticos y Computación Universidad Politécnica de Valencia.
- Kohn LT, Corrigan JM, Donaldson MS. (2000). *Institute of Medicine (US) Committee on Quality of Health Care in America To Err is Human: Building a Safer Health System*. National Academies Press (US); DOI: <https://doi.org/10.17226/9728>
- Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, Hebert L, Newhouse JP, Weiler PC, Hiatt H. (1991) *The nature of adverse events in*

*hospitalized patients. Results of the Harvard Medical Practice Study II.* N Engl J Med, 324(6):377-84. doi: <https://10.1056/NEJM199102073240605>.

Ley Federal de protección de datos personales en posesión de los particulares, DOF 05-07-2010 (2010). <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

Ley general de salud. DOF 26-03-2024 (1984) <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGS.pdf>

Lutz, M. (2013). *Learning Python: Powerful object-oriented programming* (5th ed.). O'Reilly Media.

Martin A Makary M.A., Daniel, M. (2016). *Medical error the third leading cause of death in the US.* Department of Surgery, Johns Hopkins University School of Medicine, Baltimore. [https://www.hcp.med.harvard.edu/sites/default/files/methodsmakarydaniel\\_2016\\_bmj%20\(1\).pdf](https://www.hcp.med.harvard.edu/sites/default/files/methodsmakarydaniel_2016_bmj%20(1).pdf)

Moine., J. (2013) *Metodologías para el descubrimiento de conocimiento en base de datos: un estudio comparativo.* [Tesis de magister]. Facultad de Informática de la Universidad Nacional de la Plata. <https://core.ac.uk/download/pdf/16703288.pdf>

NORMA Oficial Mexicana NOM-024-SSA3-2012, Sistemas de información de registro electrónico para la salud. Intercambio de información en salud. DOF: 30/11/2012 (2012) [https://dof.gob.mx/nota\\_detalle.php?codigo=5280847&fecha=30/11/2012#gsc.tab=0](https://dof.gob.mx/nota_detalle.php?codigo=5280847&fecha=30/11/2012#gsc.tab=0)

Organización Mundial de la Salud. [OMS]. (2019). *La OMS hace un llamamiento urgente para reducir los daños causados al paciente en la atención de salud.* <https://www.who.int/es/news/item/13-09-2019-who-calls-for-urgent-action-to-reduce-patient-harm-in-healthcare>

Organización Mundial de la Salud. (2023). *Global report on hypertension: The race against a silent killer.* [https://www.who.int/news/item/19-09-2023-launch-of-the-first-who-global-report-on-hypertension#:~:contentReference\[oaicite:0\]{index=0}&#8203;:contentReference\[oaicite:1\]{index=1}&#8203;:contentReference\[oaicite:2\]{index=2}](https://www.who.int/news/item/19-09-2023-launch-of-the-first-who-global-report-on-hypertension#:~:contentReference[oaicite:0]{index=0}&#8203;:contentReference[oaicite:1]{index=1}&#8203;:contentReference[oaicite:2]{index=2}).

- Organización Mundial de la Salud. (1994). Guía de la buena prescripción: Un enfoque práctico. <https://apps.who.int/iris/handle/10665/59001>
- Ozaydin, B. Michael-Hardin, J. Chnieng, D. (2016) *Data Mining and Clinical Decision Support Systems*. En E. S. Berner (Ed.) *Clinical Decision Support Systems* (pp. 45-68) Springer.
- Parva E, Boostani R, Ghahramani Z, Paydar S. (2017). *The Necessity of Data Mining in Clinical Emergency Medicine; A Narrative Review of the Current Literature*. Bull Emerg Trauma. 5(2), 90-95. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5406178/>
- Radicelli-Garcia, C.D., Samaniego-Erazo, G.N., Benitez-Pérez, V.E., Palacios-Trujillo, E.P. & Flor-Mora, O.P. (2021). *Estrategias de mejoramiento continuo en la gestión académica e investigativa universitaria, utilizando técnicas de minería de datos*. Universidad Nacional de Chimborazo (Unach). <https://doi.org/10.37135/u.editorial.05.33>.
- Romero-Saldivar, Y., Ramírez-Pérez, J.F., Soto-Pelegrin, L., (2022). *La minería de datos en apoyo a la toma de decisiones clínicas*. Revista Cubana de Transformación Digital, 3(2), 1-10. <http://portal.amelica.org/ameli/journal/389/3893437005/>
- Saber-Tehrani, A. Lee, H. Mathews, S. Shore, A. Makary, M. Pronovost, P. Newman-Toker, D. (2013). *25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National Practitioner Data Bank*. BMJ Qual Saf, 22(8), 672-680. DOI: 10.1136/bmjqs-2012-001550
- Santamaria-Benhumea, N.H., Garduño-Alanis, A., de la Rosa Parra, J.A., Santamaria-Benhumea, M., Rueda-Villalpando, J.P, Herrera-Villalobos, J.E (2014). *Aspectos legales del error médico en México*. REVISTA CONAMED, 19 (1), 32 –36. [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&act=8&ved=2ahUKEwj9667Att7\\_AhVLiO4BHTuYB-wQFnoECA8QAQ&url=https%3A%2F%2F Dialnet.unirioja.es%2Fdescarga%2Farti](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&act=8&ved=2ahUKEwj9667Att7_AhVLiO4BHTuYB-wQFnoECA8QAQ&url=https%3A%2F%2F Dialnet.unirioja.es%2Fdescarga%2Farti)

culo%2F4730769.pdf&usg=AOvVaw10bkHahM\_7QobWuKZWvinv&opi=8997844  
9

Secretaria de Salud. Comisión Nacional de Arbitraje Médico (2023). Anuario estadístico 2022. México, 2023.

[https://www.gob.mx/cms/uploads/attachment/file/818540/Anuario\\_2022\\_CONAMED.pdf](https://www.gob.mx/cms/uploads/attachment/file/818540/Anuario_2022_CONAMED.pdf)

Secretaria de Salud. Comisión Nacional de Arbitraje Médico (2012). NOM-024-SSA3-2012. Diario Oficial de la Federación. [https://dof.gob.mx/nota\\_detalle.php?codigo=5280847&fecha=30/11/2012](https://dof.gob.mx/nota_detalle.php?codigo=5280847&fecha=30/11/2012)

Stair R. M. Reynolds G. W. Coro Pando J. Blanco J. L. & Mercado E. (2000). *Principios de sistemas de información : enfoque administrativo* (4a ed.). Thomson.

Stallman, R. (2002). *Free software, free society: Selected essays of Richard M. Stallman*. GNU Press.

Vicente-Cestero, E., Mateos-Caballero, A (2018). *Data Science y Redes Complejas Métodos y aplicaciones*. Universitaria Ramón Areces.

Witten, I., Frank, E., Hall, M. (2011). *Data Mining Practical Machine Learning Tools and Techniques*. Editorial. <https://www.sciencedirect.com/book/9780123748560/data-mining-practical-machine-learning-tools-and-techniques>

World Health Organization. (2003). *Manual de la OMS sobre el manejo de enfermedades comunes*. WHO Press.

Zhang Y, Guo SL, Han LN, Li TL. (2016) *Application and Exploration of Big Data Mining in Clinical Medicine*. Chinese Medical Journal, 129(6), 731-737.

## Glosario

### D

DOF: Diario Oficial de la Federación.

### C

CECAMET: Comisión Estatal de Conciliación y Arbitraje Médico de Tabasco.

CONAMED: Comisión Nacional de Arbitraje Médico.

CRISP-DM: Cross Industry Standard Proces for Data Mining

### I

ISSSTE: Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado.

IA: Inteligencia Artificial.

### O

OPS: Organización internacional especializada en Salud Publica de las Américas.

OMS: Organización Mundial de la Salud.

### K

KDD Kenowledge Discovery in Databases

### R

RNA: Redes Neuronales Artificiales.

### S

SSD: Sistema de Soporte a la toma de Decisión.

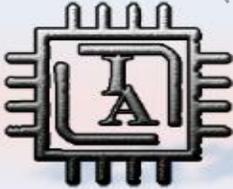
SEMMA: Sample Explore Modify Model Assess

# Apéndice A Artículo publicado en la Revista Investigación Aplicada, un Enfoque en la Tecnología

Figura

33

Carta de aceptación del artículo *Uso de la minería de datos en un sistema de apoyo a la toma de decisiones para determinar el tratamiento clínico adecuado*



**Revista Investigación Aplicada,  
un Enfoque en la Tecnología**

<https://www.investigacionaplicadarevista.com/>  
Reserva de Derechos al Uso Exclusivo No. 04-2017-061312174600-203,  
ISSN: 3594-035X, ambos otorgados por el Instituto Nacional del Derecho de Autor.

17 de diciembre de 2023

**Estimados autores:**  
Ismael Zarate Hernández  
Guillermo de los Santos Torres  
Eric Ramos Méndez

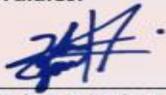
**Artículo:**

**“Uso de la minería de datos en un sistema de apoyo a la toma de decisiones para determinar el tratamiento clínico adecuado”**

Tenemos el gusto de informarles que el artículo arriba citado ha sido aprobado para su publicación en la **Revista Investigación Aplicada, un Enfoque en la Tecnología**, en la edición número **16**, correspondiente a **julio-diciembre de 2023**.

La revista es **arbitrada** y revisada en pares por evaluadores con amplia experiencia en la Investigación. Reserva de Derechos al Uso Exclusivo No. 04-2017-061312174600-203, ISSN: 3594-035X, ambos otorgados por el Instituto Nacional del Derecho de Autor. Indexada por Latindex con Folio: 28438, <https://latindex.org/latindex/ficha/28438>; ORCID Folio: 0000-0003-4920-8606, Researcher ID: AGJ-9088-2022, Academia.edu: <https://independent.academia.edu/Investigacion%20en%20Tecnologia>, Google Académico (Google Scholar): <https://scholar.google.com/citations?hl=es&authuser=2&user=ABDckAAAAJ>  
Con DOI Internacional (Digital Object Identifier) **10.60908/iaet.3594-035X**.

**Saludos cordiales.**

  
Dra. Zulma Sánchez Estrada

  
Dr. Jorge Noriega Zenteno

Comité Editorial

**IAET-DIC-23-718**

Revista Arbitrada, Indexada y DOI

Latindex. Folio: 28438  
DOI: [10.60908/iaet.3594-035X](https://doi.org/10.60908/iaet.3594-035X)  
ORCID Folio: 0000-0003-4920-8606  
Researcher ID: AGJ-9088-2022  
Academia.edu:  
<https://independent.academia.edu/Investigacion%20en%20Tecnologia>  
Google Académico (Google Scholar):  
<https://scholar.google.com/citations?hl=es&authuser=2&user=ABDckAAAAJ>

Revista Investigación Aplicada,  
un Enfoque en la Tecnología



[Investigacionaplicada4@gmail.com](mailto:Investigacionaplicada4@gmail.com)  
Tel. 5511265527

## Apéndice B Artículo publicado en la International Association for Digital Transformation and Technological Innovation

Figura

34

*Carta de aceptación del artículo Modelo de minería de datos para la detección de enfermedades en pacientes de primer nivel de atención médica*



# ICITED 24

## ACCEPTANCE & INVITATION LETTER

May 10, 2024

Dear ISMAEL ZARATE HERNANDEZ, ERIC RAMOS MENDEZ, GUILLERMO DE LOS SANTOS TORRES, GERARDO ARCEO MOHENO  
UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO  
MEXICO

On behalf of the ICITED'24 - The 4th International Conference in Information Technology & Education, I am pleased to inform that your submission "*Modelo de minería de datos para la detección de enfermedades en pacientes de primer nivel de atención médica*" has been accepted for publication in RISTI - Iberian Journal of Information Systems and Technologies Indexed in SCOPUS and presented in this conference as Full Paper. ICITED'24 to be held at the Pernambuco University, in Recife, Brazil, between the 11th and the 13th of July 2024 is a multidisciplinary conference with a special focus on new information technologies in the Education sector.

Sincerely,

João Vidal Carvalho

António José Abreu Silva

ICITED'2024, General Chairs

## Apéndice C Artículo publicado en la International Conference on Information Technology & System

**Figura**

**35**

Carta de aceptación del artículo *Development of a data mining model using the clustering technique and the A priori algorithm to determine medical treatments in second level medical care hospitals*



### ACCEPTANCE LETTER

Dear Ismael Zarate Hernandez, Guillermo de los Santos Torres, Eric Ramos Méndez, Gerardo Arceo Moheno, and Pablo Payro Campos  
Universidad Juárez Autónoma de Tabasco  
Mexico

On behalf of the ICITS'25 - The 2025 International Conference on Information Technology & Systems, I am pleased to inform you that your submission "*Development of a data mining model using the clustering technique and the Apriori algorithm to determine medical treatments in second level medical care hospitals*" has been accepted as a Full Paper for publication and oral presentation in this conference.

So, you are cordially invited to participate and present the paper in ICITS'25 (<http://www.icits.me/>) to be held in Mexico City, Mexico, between the 22th and the 24th of January of 2025, an international scientific event sponsored and organized by Politécnico Nacional, ITMA and IEEE SMC.

We sincerely hope that you will join us in making ICITS'25 a success. We look forward to seeing you next January.

Sincerely,

A handwritten signature in blue ink, appearing to read "Alvaro Manuel Reis da Rocha".

Álvaro Manuel Reis da Rocha

ICITS'25, General Chair

## Anexo A. Formato de evaluación del sistema de recomendaciones médicas

### Formato de Evaluación del Sistema de Evaluación del Sistema de Recomendaciones Médicas

El objetivo de esta evaluación es asegurar que el Sistema de Recomendaciones Médicas proporcione un servicio confiable, preciso y fácil de usar para los profesionales de la salud. Nos interesa saber cómo se comporta el sistema en la práctica real y cómo se puede optimizar para ofrecer un mejor servicio. Los resultados de esta evaluación serán utilizados para hacer ajustes técnicos, mejorar la experiencia de usuario, y garantizar que el sistema cumpla con los estándares de calidad en el ámbito médico.

**Instrucciones:** Por favor, responde a las siguientes preguntas indicando tu nivel de acuerdo con cada afirmación utilizando la escala de Likert que va de 1 a 5, donde:

- 1: Totalmente en desacuerdo
- 2: En desacuerdo
- 3: Neutral
- 4: De acuerdo
- 5: Totalmente de acuerdo

#### **Usabilidad del Sistema**

1. El sistema es fácil de usar.  
1  2  3  4  5
2. La interfaz del sistema es clara y comprensible.  
1  2  3  4  5
3. Los campos de entrada están bien organizados y son fáciles de encontrar.  
1  2  3  4  5
4. La navegación por el sistema es intuitiva.  
1  2  3  4  5
5. El sistema proporciona suficiente información de ayuda o guía cuando es necesario.  
1  2  3  4  5
6. El diseño visual del sistema es atractivo.  
1  2  3  4  5

#### **Operatividad del Sistema**

1. El sistema responde de manera rápida a las entradas del usuario.  
1  2  3  4  5

2. El botón "Predecir Medicamentos" funciona correctamente en cada intento  
1  2  3  4  5
3. Los tiempos de carga para generar las predicciones son aceptables.  
1  2  3  4  5
4. La funcionalidad del sistema es consistente y estable durante su uso.  
1  2  3  4  5
5. El sistema permite corregir entradas erróneas de manera sencilla.  
1  2  3  4  5
6. El sistema ha tenido errores o fallos durante el uso.  
1  2  3  4  5

### Resultados del Sistema

1. Las recomendaciones de medicamentos parecen adecuadas según la información proporcionada.  
1  2  3  4  5
2. El sistema genera recomendaciones basadas en los síntomas y afecciones correctamente.  
1  2  3  4  5
3. Los resultados del sistema son claros y comprensibles.  
1  2  3  4  5
4. El sistema proporciona información adicional útil, como descripciones de los medicamentos.  
1  2  3  4  5
5. Los resultados del sistema cumplen con mis expectativas.  
1  2  3  4  5

---

### Evaluación General del Sistema

En general, ¿cómo calificarías el sistema de recomendaciones médicas?

- [  ] Excelente  
[  ] Bueno  
[  ] Regular  
[  ] Malo
- 

**Comentarios adicionales:** (Escribe cualquier observación adicional sobre tu experiencia con el sistema o posibles áreas de mejora).