

TECNOLÓGICO NACIONAL DE MÉXICO INSTITUTO TECNOLÓGICO DE HERMOSILLO

SISTEMA DE DETECCIÓN Y RECONOCIMIENTO
PATRONES EN BASES DE DATOS DE ADICCIONES
UTILIZANDO LA INTELIGENCIA ARTIFICIAL

TESIS

COMO REQUISITO PARCIAL PARA OBTENER EL
GRADO DE:

MAESTRÍA EN CIENCIAS DE LA
COMPUTACIÓN

PRESENTADO POR:

ING. EDGAR GENARO ESTRADA
RIOS

DIRECTOR: DRA. MARÍA TRINIDAD SERNA
ENCINAS





EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Hermosillo

"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

SECCIÓN: DIV. EST. POS. E INV.
No. OFICIO: DEPI/077/20
ASUNTO: AUTORIZACIÓN DE IMPRESIÓN DE TESIS.

10 de julio de 2020

**C. EDGAR GENARO ESTRADA RÍOS,
PRESENTE.**

Por este conducto, y en virtud de haber concluido la revisión del trabajo de tesis que lleva por nombre **"SISTEMA DE DETECCIÓN Y RECONOCIMIENTO DE PATRONES EN BASES DE DATOS DE ADICCIONES UTILIZANDO LA INTELIGENCIA ARTIFICIAL"**, que presenta para el examen de grado de la MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN, y habiéndola encontrado satisfactoria, nos permitimos comunicarle que se autoriza la impresión del mismo a efecto de que proceda el trámite de obtención de grado.

Deseándole éxito en su vida profesional, quedo de usted.

ATENTAMENTE

M.C. MARÍA TRINIDAD SERNA ENCINAS
DIRECTORA

M.C. CÉSAR ENRIQUE ROSE GÓMEZ
SECRETARIO

M.C. RAFAEL ARMANDO GALAZ BUSTAMANTE
VOCAL

M.C. ROSA IRENE SÁNCHEZ FERMÍN
JEFA DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN



S.E.P.

INSTITUTO TECNOLÓGICO
DE HERMOSILLO
DIVISIÓN DE ESTUDIOS
DE POSGRADO

RISF/fjcg*



Av. Tecnológico S/N Col. El Sahuaro C.P. 83170 Hermosillo, Sonora
Tel. 01 (662) 2-606500, Ext. 136 e-mail: depi_hermosillo@tecnm.mx
www.tecnm.mx | www.ith.mx



ISO 9001:2015
Sistema de Gestión de Calidad Certificado

Resumen

El consumo de sustancias psicoactivas presenta un gran reto a nivel mundial, estudios indican que año con año se ha presentado un incremento en el uso de estas. En México el consumo de alcohol ha tenido un incremento considerable con el pasó del tiempo, mientras que Sonora se encuentra entre los 10 estados con mayor consumo de esta sustancia.

Como parte de uno de los programas implementados en el estado de Sonora se encuentra el Sistema de Vigilancia Epidemiológica de las Adicciones (SISVEA), el cual realiza evaluación, tratamiento y seguimiento de las personas con este tipo de problema, si bien, este tipo de programas están conducidos por expertos en el tema, no cuentan con herramientas tecnológicas que apoyen la carga al realizar el proceso de evaluación. Es por ello, que surge la necesidad de implementar herramientas informáticas como soporte para este proceso.

En esta investigación se presenta la propuesta, diseño, implementación y pruebas realizadas de un algoritmo de inteligencia artificial para el reconocimiento de patrones en bases de datos de adicciones.

Para alcanzar el objetivo del sistema, se utilizó el algoritmo de inteligencia artificial llamado máquina de soporte vectorial (SVM), el cual fue comparado con diferentes algoritmo, y obtuvo un resultado sobresalientes ante los demás, arrojando un 76 % de precisión.

Índice general

Índice de figuras	7
1. Introducción	1
1.1. Antecedentes	3
1.2. Planteamiento del Problema	5
1.3. Objetivos	7
1.3.1. Objetivo general	7
1.3.2. Objetivos específicos	7
1.4. Justificación	8
1.5. Alcances y delimitaciones	8
1.6. Metodología	8
1.7. Organización de la tesis	9
2. Marco teórico	11
2.1. Introducción	11
2.2. Adicciones	11
2.2.1. ¿Cómo se clasifican las adicciones?	12
2.2.2. ¿Cuáles son las etapas de las adicciones?	12
2.3. Drogas	14
2.3.1. Clasificación de las Drogas	14
2.4. Violencia	18
2.5. Big Data	19
2.6. Inteligencia Artificial	22
2.6.1. Algoritmos Predictivos	23

2.7.	Trabajos Relacionados	30
2.7.1.	AutoOPU	30
2.7.2.	Aprendizaje automático como un complemento de la toma de decisiones clínicas en el tratamien- to de dependencia del alcohol	31
2.7.3.	NeuCube	31
2.7.4.	Triggr Health	32
2.7.5.	Predicción del uso de sustancias en los medios sociales	32
3.	Análisis y diseño del sistema	35
3.1.	Introducción	35
3.2.	Análisis del sistema	35
3.2.1.	Diagramas de contexto: Nivel 0	35
3.2.2.	Diagrama de contexto: Nivel 1	36
3.2.3.	Casos de uso	37
3.3.	Diseño del Sistema	44
3.3.1.	Diagrama de Actividades	44
3.3.2.	Diagrama de Clases	48
3.3.3.	Diagrama de Componentes y Despliegue	49
3.4.	Diagrama de base de datos	50
3.5.	Arquitectura propuesta del sistema	52
4.	Implementación	53
4.1.	Recolección de los datos	53
4.2.	Análisis de atributos	56
4.3.	Experimentación	58
4.3.1.	Primera experimentación con datos de 2017	63
4.3.2.	Segunda experimentación con datos de 2017-2018	64
4.3.3.	Tercera experimentación con datos de 2014-2018	65
4.3.4.	Cuarta experimentación con datos de 2014-2018	66
4.3.5.	Quinta experimentación con datos de 2014-2018	66
4.3.6.	Sexta experimentación con datos de 2014-2018	67
4.3.7.	Séptima experimentación con datos de 2014-2018	68

4.3.8. Octava experimentación con datos de 2014-2019	68
4.3.9. Novena experimentación con datos de 2014-2019	69
4.3.10. Validación del algoritmo utilizado	69
4.4. Capturas del Sistema	73
5. Análisis de resultados	83
5.1. Primeros Resultados	83
5.2. Resultados utilizando validación cruzada	85
6. Conclusiones y trabajo futuro	89
6.1. Conclusiones	89
6.2. Trabajo futuro	90
Bibliografía	93

Índice de figuras

- 2.1. Imagen de Heroína en sus diferentes presentaciones [1]. 15
- 2.2. Era actual de la inteligencia artificial [2]. 20
- 2.3. Arquitectura de una red neuronal. 26
- 2.4. Datos linealmente separables. 28

- 3.1. Diagrama de contexto: nivel 0. 36
- 3.2. Diagrama de contexto: nivel 1. 37
- 3.3. Diagrama de casos de uso del sistema. 38
- 3.4. Diagrama de secuencia de caso de uso Login. 39
- 3.5. Diagrama de secuencia de caso de uso Tamizaje. 40
- 3.6. Diagrama de secuencia de caso de uso Catálogos. 41
- 3.7. Diagrama de secuencia de caso de uso Catálogo extendido agregar. 42
- 3.8. Diagrama de secuencia de caso de uso Catálogo extendido editar y/o eliminar. 42
- 3.9. Diagrama de secuencia de caso de uso Preprocesamiento. 43
- 3.10. Diagrama de secuencia de caso de uso Implementación de algoritmo. 43
- 3.11. Diagrama de secuencia de caso de uso Análisis de resultados. 44
- 3.12. Diagrama de actividad para Caso de uso Login. 45
- 3.13. Diagrama de actividad para Caso de uso Tamizaje. 46
- 3.14. Diagrama de actividad para Caso de uso Catálogos. 47

3.15. Diagrama de actividad para Caso de uso Análisis de resultados	48
3.16. Diagrama de Clases	49
3.17. Diagrama de Componentes y despliegue.	50
3.18. Diagrama de la Base de Datos.	51
3.19. Diagrama de Arquitectura propuesta del sistema.	52
4.6. Algoritmo de lectura de <i>“dataset”</i>	58
4.7. Algoritmo de configuración de métricas.	60
4.8. Algoritmo de asignación y de ejecución.	61
4.9. Gráfica de resultados en comparativa de los algoritmos.	62
4.11. Líneas de código actualizadas.	65
4.12. Iris Setosa.	70
4.13. Experimentación con Orange y WEKA.	71
4.14. Código de ejecución de la validación cruzada.	72
4.15. Captura de la pantalla de login.	73
4.16. Captura de la pantalla principal.	74
4.17. Captura del módulo de centro.	75
4.18. Captura de la pantalla para agregar un centro.	76
4.1. Cara frontal del formato de tamizaje de los años 2014-2016.	77
4.2. Cara trasera de formato del tamizaje de los años 2014-2016.	78
4.3. Cara frontal del formato de tamizaje de los años 2017-2019.	79
4.4. Cara trasera del formato de tamizaje de los años 2017-2019.	80
4.5. Formatos del tamizaje de los años A) 2014-2016 y B) 2017-2018.	81
4.10. Código de ejecución de las experimentaciones.	82

Capítulo 1

Introducción

La adicción, a una o más sustancias, es uno de los principales problemas que enfrentan los servicios de salud a nivel mundial, de acuerdo con la Oficina de Naciones Unidas contra las Drogas y el Delito (UNODC), se calcula que alrededor de 271 millones de personas entre 15 y 64 años, consumieron algún tipo de droga en 2017; esta estimación revela un incremento en un 30 % en el uso de sustancias de este tipo con respecto al año 2009 [3]. Por otra parte, según la organización mundial de la salud (OMS), el consumo total de alcohol per cápita en la población mundial mayor de 15 años, subió de 5.5 litros de alcohol puro en 2005 a 6.4 litros en 2010, manteniéndose este nivel de 6.4 litros en 2016. Debido al consumo desmedido de esta sustancia, se estima que en 2016 provocó aproximadamente 3 millones de muertes, lo cual equivale al 5.3 % de las defunciones totales [4].

En algunos países incluyendo México, se ha incrementado el uso simultáneo o sucesivo de combinaciones de sustancias, lo que agrava los efectos tóxicos que éstas puedan causar; de manera proporcional, también se suele modificar la vía de administración, esto con el fin de acentuar el efecto que éstas provocan [5].

De acuerdo a la Comisión Nacional contra las Adicciones (CONA-

DIC) en su reporte 2016-2017, en México el porcentaje de consumo de drogas aún es bajo en comparación con otros países [6]; sin embargo, este porcentaje ha tenido un incremento año tras año; mientras que el consumo de alcohol ha tenido un incremento considerable, especialmente en el patrón de consumo excesivo, y es en las mujeres donde este porcentaje ha tenido un mayor incremento. El tabaco no representa una gran diferencia en comparación con el consumo de alcohol, si bien el número de consumidores es menor que el del alcohol, el tabaco ha tenido un importante crecimiento. Estos números muestran que el 17.6% de la población de 12 a 65 años fuma tabaco actualmente, lo cual corresponde a 14.9 millones de fumadores mexicanos, entre los cuales el 6.4% fuma diariamente (5.5 millones) y el 11.1 fuma de manera ocasional (9.4 millones) [7].

Específicamente en el estado de Sonora, el porcentaje de uso de algún tipo de droga se ha duplicado en un lapso de 8 años (2008-2016); sin embargo, Sonora se encuentra entre los 10 estados que más alcohol consumen [8]; en cuanto a tabaco, también se encuentra dentro de la media nacional con un 17%, donde un 59% intentó dejar de fumar sin éxito [7].

La existencia de una adicción no sólo se presenta en personas adultas, este problema suele iniciar en el rango de edades cada vez más tempranas, esto sucede especialmente en el consumo de alcohol y cannabis [9][10]. Gracias a la gran aceptación social que éstos tienen, cada vez se vuelve más común el ver jóvenes en estado de ebriedad e incluso esto llega a fomentarse en el núcleo familiar. En algunos casos, al percibirse a temprana edad, se le considera normal y su percepción de riesgo es baja, sin embargo, ésta es solo una de las variables sociales, familiares, personales y cognitivas [11].

Alrededor del mundo se busca mejorar los servicios de salud, por lo cual se están implementando algoritmos computacionales, que permitan mejorar la eficiencia y administración de los servicios de salud.

En base a esto, en esta investigación se desarrollará un sistema para el reconocimiento de patrones y/o la predicción del patrón de consumo, utilizando la inteligencia artificial, lo que permitirá una atención temprana en los casos detectados.

1.1. Antecedentes

Las drogas han acompañado al hombre en su evolución, es de suponer que la primera droga que el hombre utilizó, fuese el resultado de la fermentación de algún fruto, por lo que es evidente que para que esto ocurriera, fue necesario la invención de algún tipo de recipiente, lo cual ayuda a aproximar la fecha de su creación. Las primeras suposiciones sobre la existencia de alcohol se cifran en el quinto milenio AC. Se sabe con una alta probabilidad, que el aguamiel fue el primer alcohol que el humano produjo y consumió. Fue por los efectos psicotrópicos que éstas causaron, que se reservó su uso con fines religiosos [12].

La curiosidad del hombre lo llevó a experimentar con un gran número de especies vegetales, hasta que esa curiosidad lo llevó al descubrimiento del tabaco, no se sabe con exactitud si ésta fue la segunda droga descubierta por el hombre, pero sí se sabe que fue ésta la que tuvo una gran propagación a lo largo de mundo, a pesar de las prohibiciones que tenían un gran número de países; esto nos da a una idea de lo intenso de la dependencia que éste produce [13].

En su proceso por descubrir nuevas sustancias que provocaran efectos psicotrópicos, el hombre se percató de que ciertas hierbas podían apaciguar dolores, con lo cual se descubren los analgésicos, ésta es la primer referencia que se tiene del uso de opio; es en base a esto, que el hombre pudo catalogar el uso de estas sustancias como médicas o religiosas, dependiendo del uso que se le daba [14].

Si bien la existencia de las drogas nos ha acompañado en toda nuestra evolución, e incluso éstas han ayudado al hombre en su desarrollo, es necesario delimitar el uso que se le da a éstas. Para ello es necesario saber la diferencia entre uso y abuso de drogas, donde el uso se refiere a la utilización esporádica de la sustancia sin algún fin médico, mientras que el abuso se define como un consumo de manera prolongada, el cual incluso puede producir cambios en el funcionamiento del cerebro y llevarlo a un desorden permanente del comportamiento. Si bien el uso y abuso de drogas se puede evitar, existe un tercer ámbito, la drogodependencia o farmacodependencia, éste es el caso donde se consume de manera crónica, con el fin de experimentar sus efectos psíquicos y en ocasiones, para evitar el malestar producido por la privación de la sustancia [15].

Sin embargo, la explicación a el consumo de manera abusiva de las drogas se encuentra en lo que se ha denominado «Teoría de la espiral de desregulación del sistema de recompensas cerebrales», según la cual, la adicción se inicia desde el primer contacto con la droga e implica el inicio de un proceso adaptativo que el organismo desarrolla para contrarrestar los efectos que ésta tiene, en especial a nivel del sistema nervioso, debido a que el organismo siempre intenta mantener un equilibrio homeostático en todos sus sistemas, incluyendo el sistema de recompensa cerebral. El uso de una droga de manera abusiva y la aparición de sus efectos, desequilibra dicha homeostasis, lo que desencadena el desarrollo de una respuesta adaptativa, donde el cuerpo intenta combatir el efecto inicial de la droga. Debido a esta reacción adaptativa, los circuitos cerebrales implicados en el refuerzo no vuelven al nivel fisiológico que mantenían antes de la exposición a la droga [16]. Actualmente, la integración de sistemas inteligentes al sector salud ha ayudado en gran medida al diagnóstico médico, esto gracias a sus posibilidades para involucrarse en situaciones que requieren un gran acervo de conocimientos médicos, rapidez en el procesamiento de datos y la toma efectiva de decisiones, que estos sistemas poseen.

Los sistemas inteligentes pretenden emular la capacidad natural del hombre para la toma de decisiones, imitando desde la manera en que éste aprende, hasta la forma de reaccionar, dependiendo de la situación que se le presente. [17], es por ello que podemos decir que la IA busca entender la manera en que el humano percibe, comprende y soluciona problemas, con el fin de replicar este comportamiento mediante una máquina [18].

La inteligencia artificial (IA) en relación al tema de adicciones no ha tenido una gran participación, esto se debe a la alta complejidad que este problema presenta; sin embargo, se han presentado aplicaciones las cuales, mediante el procesamiento del lenguaje natural y la aplicación de Machine Learning, logran detectar pacientes que reincidieron en este problema o que están en etapa de descubrimiento de la droga en cuestión [19].

1.2. Planteamiento del Problema

Actualmente, las adicciones a cualquier tipo de droga legal o ilegal es un problema de salud complicado, este problema implica varias categorías y niveles, desde su elaboración y distribución, hasta sus efectos después del consumo [20]; dicho problema es cíclico, puesto que el mismo consumo crea una necesidad, sin mencionar que en la mayoría de los casos, el efecto que éstas causan, deja de ser suficiente para la persona y se tiende a incrementar la dosis o incluso escalar a una sustancia con mayor concentración y riesgo [21].

Como parte de la complejidad de este problema, cabe mencionar la interrelación que existe entre las adicciones y la violencia, puesto que es común que una persona adicta, con el fin de conseguir el dinero necesario para adquirir la sustancia o sustancias, a la que la persona reacciona actualmente, esté dispuesta a realizar actos ilícitos poniendo

en riesgo su seguridad, la seguridad de alguien más o los bienes materiales [10].

Otra relación a mencionar en el uso desmedido de las sustancias legales o ilegales, es el deterioro de la salud; en el caso de las mujeres, éstas pueden estar dispuestas a consumir algún tipo de sustancia aún en condiciones de embarazo, provocando problemas en el desarrollo del feto; así también, las personas que se inyectan cualquier tipo de sustancia, sin utilizar una jeringa esterilizada, se exponen a contraer una enfermedad como el VIH, ayudando así a la propagación de éste. Según la OMS, a nivel mundial hay alrededor de 16 millones de personas que se inyectan drogas, y de ellas, aproximadamente 3 millones están infectadas por el VIH. En promedio, una de cada diez infecciones nuevas por VIH es causada por el consumo de drogas inyectables [22].

Actualmente, se gastan alrededor de 29 mil millones de pesos en salud por enfermedades asociadas sólo al tabaquismo. La tolerancia social frente al consumo de tabaco, tanto en hombres como en mujeres, favorece la adicción a esta sustancia [23].

Como resultado de lo anterior, el presente trabajo plantea las siguientes preguntas de investigación:

¿Qué es una adicción y sus tipos?

¿Cuales bases de datos tienen información sobre adicciones?

¿Es necesario integrar la información disponible?

¿Qué algoritmos de IA pueden aplicarse para el reconocimiento de patrones en las adicciones?

¿Qué tipo de predicciones se puede realizar con la información dis-

ponible?

Con esto surge la siguiente pregunta que guía esta investigación: En relación con las Adicciones, ¿Cuáles funcionalidades debe tener un sistema inteligente para prevenir y/o predecir tendencias en las adicciones, que permita una detección oportuna y, en consecuencia, una atención temprana en los casos detectados?

1.3. Objetivos

Para el desarrollo de la presente investigación, se identificaron los siguientes objetivos.

1.3.1. Objetivo general

Implementar un sistema inteligente para la detección y reconocimiento de patrones, que identifique comportamientos conductuales en pacientes que presentan dependencia a una o varias drogas, lo que permitirá prevenir y/o predecir tendencias hacia alguna adicción.

1.3.2. Objetivos específicos

- Conocer la información existente en las bases de datos de adicciones.
- Integrar y pre-procesar los datos requeridos.
- Analizar y seleccionar los algoritmos de IA a utilizar.
- Implementar el sistema inteligente de reconocimiento de patrones y/o análisis predictivo.
- Analizar y validar los resultados obtenidos.

1.4. Justificación

Las adicciones son un problema de salud pública, que perjudican no sólo a la persona que la padece, sino también a la sociedad en general, ayudando a dispersar este consumo y causando un incremento de personas con este problema o con la posibilidad de ser un riesgo para la sociedad. Por lo anterior, se tiene interés en aplicar tecnologías de inteligencia artificial, que permitan mejorar y eficientar los recursos que disponen los servicios de salud del estado de Sonora. Aplicar la IA en los modelos de salud, sienta las bases para la creación de herramientas que posteriormente ayudarán a los pacientes a detener su desarrollo en las adicciones, previniendo un deterioro mayor en su salud.

1.5. Alcances y delimitaciones

El sistema utilizará los datos contenidos en los tamizajes referente a las personas con alguna adicción, atendidas en los centros de salud del estado de Sonora. El desarrollo del sistema, se llevará a cabo en el laboratorio de sistemas inteligentes de la Maestría en Ciencias de la Computación, perteneciente a la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Hermosillo.

1.6. Metodología

La metodología a implementar en el desarrollo del proyecto se compone de 3 fases: La primera se refiere a los fundamentos teóricos del tema, la segunda se enfoca en el manejo de datos y el diseño del modelo y una última fase donde se evalúan los resultados obtenidos.

La fase uno se enfoca en la recopilación de la literatura sobre adicciones, que puedan proveer información relevante al tema, así como el

estado actual de los algoritmos de IA, para seleccionar los más adecuados a utilizar con respecto a la información disponible.

La fase dos se compone de la recolección e integración de los datos disponibles, para generar una propuesta de solución adecuada, ésta también contempla la creación de diagramas del análisis y diseño, el modelo de datos y la esquematización de la arquitectura de software propuesta del sistema. Los cuales definirán la funcionalidad y la operación del proyecto final.

Finalmente, la fase tres se compone de la implementación del sistema inteligente, de la interpretación de los resultados de los algoritmos aplicados y del análisis y validación de los mismos.

1.7. Organización de la tesis

El capítulo 2 presenta el estado del arte, realizando un análisis de los temas principales utilizados para fundamentar esta investigación. Entre los cuales se encuentran la categorización de las drogas legales e ilegales, algoritmos de inteligencia artificial y los trabajos relacionados con nuestra propuesta.

El capítulo 3 describe el análisis y diseño del sistema. Se esquematizan los procesos del sistema y se presentan los diferentes diagramas detectados.

El capítulo 4 describe a detalle la integración de los datos utilizados en cada experimentación.

El Capítulo 5 presenta los resultados obtenidos y se analiza cada una de las experimentaciones realizadas.

Finalmente, el capítulo 6 presenta las conclusiones obtenidas por el presente trabajo de investigación y las recomendaciones a seguir para trabajos futuros.

Capítulo 2

Marco teórico

2.1. Introducción

El marco teórico de la presente investigación, proporcionará al lector una idea más clara acerca del tema. Los tópicos centrales que enmarcan este trabajo son: Adicciones, Big Data y la Inteligencia Artificial.

2.2. Adicciones

La Organización Mundial de la Salud (OMS) define adicción como una “enfermedad física y psico-emocional, que crea una dependencia o necesidad hacia una sustancia, actividad o relación”. También se define como un conjunto de signos y síntomas en los que intervienen factores psicológicos y sociales. Es una enfermedad progresiva y puede ser fatal si no se trata a tiempo, se caracteriza principalmente por episodios continuos de descontrol, distorsiones del pensamiento y negación ante la enfermedad; dicho comportamiento suele estar acompañado por muestras de decisión a obtener las sustancias psicoactivas por cualquier medio [24][25].

2.2.1. ¿Cómo se clasifican las adicciones?

Actualmente, las adicciones se clasifican en las siguientes categorías [26]:

- Eufóricas: Opio y sus derivados, así como cocaína.
- Fantásticas: Mezcalina, marihuana y beleño, entre otras.
- Embriagantes: Alcohol, éter, cloroformo y bencina.
- Hipnóticas: Barbitúricos y otros somníferos.
- Excitantes: Cafeína y tabaco, entre otras.

Además de esta clasificación, también existen “nuevas drogas o drogas de diseño”, creadas con fines recreativos. Se supone que las personas que presentan por lo menos 3 de los siguientes puntos, pueden ser catalogados como adictos [27]:

- Fuerte deseo o necesidad de consumir la sustancia.
- Dificultades para controlar su consumo.
- Malestares al interrumpir o reducir el consumo.
- Continuar su consumo a pesar de percibir de forma clara sus efectos perjudiciales.

2.2.2. ¿Cuáles son las etapas de las adicciones?

Las adicciones cuentan con 4 etapas [28]:

- Experimentación:
Es el caso donde la persona aprende de los efectos que causan estas sustancias, es aquí donde la droga sólo proporciona placer sin presentar los efectos adversos que ésta puede provocar, posteriormente, el consumo puede continuar o interrumpirse.

- **Uso:**
El compromiso con la droga es bajo, se buscan los efectos de ésta ocasionalmente, pero con episodios de cierto sentido de culpa, se inicia el consumo individual pero no existe deterioro laboral, social o familiar, y no presenta episodios de intoxicación. Sin embargo, toda droga genera dependencia física o psíquica progresivamente y es fácil caer en el abuso.

- **Abuso:**
Se llega a tener una obsesión por la sustancia y su consumo es repetido por semana, con tendencia a un aumento en la dosis, además, se presenta deterioro en la relación de familia y amigos.

- **Adicción:**
Se presentan dificultades académicas y laborales. La búsqueda de la droga se realiza de forma compulsiva, la cual obliga a seguir con el consumo para cortar el malestar que se presenta por su privación, debido a que ya existe un compromiso orgánico e incluso hay conductas de riesgo, como uso de drogas intravenosas o combinación de varias drogas.

Los síntomas físicos que normalmente se presentan por la abstinencia, es decir, por la falta de uso de la sustancia que se consumía frecuentemente, varían según la droga que se deje de utilizar. Generalmente, causan el efecto contrario a su uso, por ejemplo, en el caso del opio, el cual tiene la capacidad de anestesiarse el cuerpo, al dejar de consumirlo éste genera dolor como principal consecuencia de la privación de su uso. Mientras que de manera psicológica, el síndrome se refleja como angustia, irritabilidad o sensación de vacío, estos síntomas suelen variar por individuo. La duración del síndrome de abstinencia también varía en base a la sustancia que es utilizada [29].

2.3. Drogas

El término droga se utiliza para toda aquella sustancia psicoactiva, que al ser ingerida, afecta los procesos mentales como la cognición o la afectividad y cuya distribución está prohibida por la ley o bien restringida a su uso médico y/o farmacéutico[25].

2.3.1. Clasificación de las Drogas

En la actualidad no se tiene una definición que logre satisfacer la gran cantidad de drogas existentes, sin embargo, existe una clasificación que por su utilidad práctica es utilizada en gran medida, ésta es la clasificación realizada por el NIDA(“*National Institute on Drug Abuse*”), el cual clasifica a las drogas por su efecto estimulante o depresor [15] [30] [31].

2.3.1.1. Drogas Depresoras

Las drogas en esta categoría, son aquellas que están relacionadas directamente con las funciones del sistema nervioso, creando interferencias entre las sinapsis e incluso disminuyendo el número de conexiones entre éstas. Estos efectos dependen de la sustancia utilizada, las drogas que comúnmente se consumen son [31]:

- Alcohol:
Esta droga es la que mayor consumo tiene a nivel mundial, esto se debe a que incluso los gobiernos permiten su consumo de manera recreativa. Esta droga se puede obtener de 2 maneras, ya sea mediante la fermentación de algún fruto o por destilación, el cual es un método artificial para aumentar el grado de alcohol de una bebida. Al poco tiempo de ser ingerido, éste puede causar desorientación, deshidratación, vértigo, náuseas, descoordinación motora, entre otros [32].

- **Morfina:**
Es conocida como el fármaco que con mayor eficacia puede erradicar el dolor intenso, causado prácticamente por cualquier enfermedad, éste es generalmente usado para contrarrestar el dolor postoperatorio, tratamiento de infarto, entre otros. Es por ello que comúnmente es utilizada con fines médicos. Si bien su uso en la rama médica no está asociada a la generación de una adicción, se han presentado casos donde se desarrolla una tolerancia a ésta. Esta sustancia es un derivado del opio [33] [34].

- **Heroína:**
Es una droga semi-sintética y se deriva del opio, esta droga es aproximadamente 5 veces más fuerte que la morfina y su forma de consumo puede variar entre ser inhalada, fumada o de manera intravenosa. Los efectos que ésta provoca aparecen en un corto lapso de tiempo, estos efectos incluyen aumento en la energía, efectos analgésicos y supresión del hambre, también es posible que aparezcan efectos de somnolencia entre otros [35].



Figura 2.1: Imagen de Heroína en sus diferentes presentaciones [1].

2.3.1.2. Drogas Estimulantes

Las drogas que pertenecen a la familia de estimulantes, tienen la peculiaridad de aumentar los niveles de alerta y de atención del consumidor, éstas afectan tanto la motricidad como la cognitividad y sus principales efectos son la desinhibición, irritabilidad, agresividad, disminución de fatiga o sueño, un menor control emocional, etc. Al igual como sucede con el resto de las drogas, sus efectos dependen de la cantidad ingerida, las drogas de mayor consumo en esta categoría son [36]:

- **Tabaco:**

El tabaco es una planta que contiene nicotina y diversos carcinógenos, estas sustancias tóxicas se ven potencializadas en el proceso de fabricación y adecuación para la venta, normalmente en forma de cigarrillos, dichas toxinas son potentes y fácilmente absorbidas por el cuerpo. Si bien la nicotina es el agente adictivo, el coctel tóxico que lo acompaña optimiza los efectos de ésta [37].

- **Cocaína:**

Es un alcaloide que se obtiene de las hojas de la planta llamada coca, esta droga es un potente estimulante del sistema nervioso, que produce euforia o insomnio; los efectos de ésta tardan de 1 a 3 minutos en aparecer, y su duración puede variar entre 20 y 40 minutos, esta droga puede ser ingerida vía oral y frecuentemente su ingesta es combinada con alcohol. Los consumidores de esta droga tienden a utilizar otros medios para su consumo, como inyección intravenosa o ingerida, mediante la inhalación de los vapores emitidos al ser calentado a través de una pipa, el procedimiento anterior aumenta la potencia de este alcaloide [25].

- **Crack:**

Es el resultado de calentar las sales de la cocaína con bicarbonato, al ser ingerido, éste tarda de 3 a 6 segundos en manifestar

sus efectos, los cuales se presentan como una sensación inmediata de euforia o la desaparición de la ansiedad, una alteración del juicio, etc. Los efectos placenteros duran aproximadamente 6 minutos, tras los cuales, el consumidor se siente obligado a repetir la dosis para recuperar el estado de excitación y euforia [25].

- **Anfetaminas:**

Su componente base se encuentra en algunas especies de plantas como la ephedra y catha, los extractos de estas plantas son usados como psicoestimulantes. El uso de esta sustancia afecta el sistema central nervioso provocando un incremento en la concentración de cortisol y adrenalina, esto a su vez crea euforia, pupilas dilatadas, hipertensión en algunos casos, excitación, paranoia, etc. Estos síntomas si bien son causados de forma natural, suelen incrementarse al ser tratados con diferentes químicos, agravando también la adicción a ésta [38].

2.3.1.3. Drogas Alucinógenas/Psicodélicas

La característica principal de este tipo de drogas es que pueden causar alucinaciones, ya sea con una percepción de un objeto inexistente o la ausencia de dicho objeto aún estando presente, hipersensibilidad sensorial, deformación de la percepción corporal propia o ajena, etc. Alguna de las drogas más utilizadas que encajan en esta categoría son [36]:

- **Mariguana:**

Su origen proviene de una planta, la cual tiene como componente principal el Tetrahidrocannabinol (THC), este elemento es el que genera el efecto psicoactivo, el cual, como otras drogas, puede causar intoxicación, euforia, ansiedad, dependencia, síndrome de abstinencia, etc. Ésta es la droga ilegal de mayor consumo [39].

- **LSD:**

Es un producto semi-sintético de un derivado natural del hongo

rye, el cual tiene una extraordinaria potencia, pues sus efectos se hacen sentir con dosis muy pequeñas. Esta sustancia se sintetiza generalmente de manera ilegal, esto debido a que puede eliminarse fácilmente del organismo sin dejar rastro de su uso, esta sustancia tiene una peculiaridad, pues a pesar de que afecta el sistema central nervioso y que puede causar alucinaciones, pérdida de la razón, crisis de ansiedad, etc, no suele producir conductas de dependencia al usar cantidades pequeñas o moderadas [40] [41].

- **Éxtasis:**

Es una droga sintética tóxica para el sistema nervioso central, el uso de esta sustancia puede provocar incremento en la frecuencia cardíaca, náuseas, sudoración, insomnio, aumento en el estado de alerta, facilita la comunicación, hipersensibilidad, paranoia, ataques de pánico, etc. Tras su ingesta, sus efectos se manifiestan entre 60 y 90 minutos y se mantienen durante 3 a 5 horas, No se tiene evidencia de que origine un problema de dependencia, incluso bajo una ingesta constante; los efectos placenteros presentan una disminución mientras que los efectos desagradables presentan un incremento [42].

2.4. Violencia

El consumo de drogas ha incluido algunos tipos de violencia como variables a considerar, suele asociarse con la violencia delictiva y las muertes por accidentes; estas categorías incluyen [43]:

- **Intrafamiliar:**

Este tipo de violencia es aquella que sucede dentro de la vivienda, independientemente si el agresor reside o ha residido en este hogar. Actualmente, las mujeres son más propensas a sufrir este tipo de violencia, sin embargo, los hijos se ven afectados en gran

medida, incluso sin ser ellos la víctima física, pues al ser testigos del conflicto que existe en su hogar, su salud mental puede verse directamente afectada, con consecuencias como desarrollar un autoestima baja, inseguridad y en algunos casos, desarrollar rasgos violentos como parte de su personalidad [44] [45].

- **Auto infligida:**
Consiste en la alteración o destrucción deliberada de alguna parte del cuerpo, sin una intención suicida consciente, se dice que ésta es una forma de auto ayuda, ya que alivia temporalmente síntomas de la ansiedad abrumadora, la depresión y la despersonalización [46].
- **Interpersonal:**
Es la alteración que realiza un individuo hacia otro, la cual puede resultar en lesiones y/o daños emocionales [47].
- **Colectiva:**
Comportamiento desarrollado por un conjunto de individuos en relación a una persona e incluso otro grupo, y que generan resultados como lesiones y/o daños emocionales [47].

La OMS define a la violencia como “el uso deliberado de la fuerza física o el poder, ya sea en grado de amenaza o efectivo, contra uno mismo, otra persona o un grupo o comunidad, que cause o tenga muchas probabilidad de causar lesiones, muerte, daños psicológicos, trastornos del desarrollo o privaciones” [48].

2.5. Big Data

Big Data permite a las organizaciones almacenar, administrar y manipular grandes cantidades de datos a la velocidad correcta y en el momento adecuado para obtener los conocimientos adecuados [49], aunque el concepto de Big Data es relativamente nuevo, se podría

decir que existe desde hace tiempo, pues la acción de recopilar y almacenar grandes volúmenes de datos, para su posterior análisis, se realiza desde hace años.

El Big Data no funciona de manera independiente, sino que utiliza otros campos de la ciencia tal como se muestra en la figura 2.2.

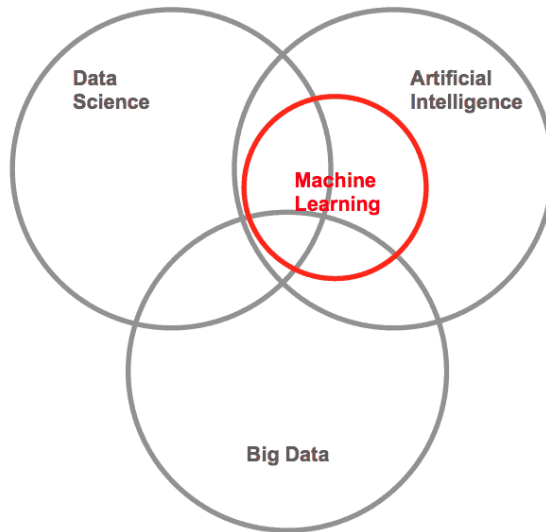


Figura 2.2: Era actual de la inteligencia artificial [2].

Éste trabaja en conjunto con otras disciplinas, como lo son la inteligencia artificial y el “*Data Science*”, esto con el fin de cumplir con el objetivo de las 5vs que lo conforman.

¿Que son las 5Vs del Big Data?

El Big Data se puede definir en base a las siguientes características: Volumen: Esta característica se enfoca en la forma en que se recolectan los datos de manera masiva y el crecimiento exponencial de los mismos. El proceso de Extracción-Integración de los datos, a partir de las bases de datos operativas de la empresa, presenta el reto de optimizar tanto el almacenamiento de los datos como el procesamiento de los mismos [50].

Variedad: Se refiere al origen de los datos, los cuales pueden variar desde los estructurados como las bases de datos y las hojas de cálculo csv; los semiestructurados como los documentos xml o páginas web; y los no estructurados como documentos de texto, audio o imágenes.

Veracidad: Debido a la amplia variedad de datos y que éstos provienen de diferentes fuentes, mantener la veracidad de los datos se vuelve un reto; es decir, se puede tener una gran cantidad de datos, pero no todos pueden ser verdaderos. Para minimizar problemas en la veracidad es necesario asegurar la confianza, autenticidad, origen y confiabilidad de los datos.

Velocidad: Una de las características principales del Big Data es reflejar un manejo y procesamiento de los datos en tiempo casi real.

Valor: Extracción de información oculta y relevante a partir del análisis de los grandes volúmenes de datos previamente adquiridos, y la capacidad de transformar el conocimiento adquirido en una mejora a la toma de decisiones [50].

Un ejemplo de aplicación de Big Data es en el campo de la medicina, puesto que podría determinar la causa de una enfermedad y brindar orientación a un médico sobre las opciones de tratamiento.

2.6. Inteligencia Artificial

“La Inteligencia Artificial (IA) es un campo de la ciencia y la ingeniería que se encarga de la comprensión, desde el punto de vista informático, de lo que se denomina comúnmente comportamiento inteligente” [51]; es decir, la IA tiene como objetivo el entender de qué manera el humano comprende, percibe y soluciona problemas con el fin de replicar dicho comportamiento mediante una computadora [18]; sin embargo, aún no se tiene una definición exacta de lo que es la inteligencia artificial.

Para definir si una máquina es inteligente o no, se utiliza una prueba llamada “*Turing Test*”, la cual fue planteada en 1950 y aún se utiliza en la actualidad, en esta prueba se lleva a cabo una conversación entre un humano y una máquina, si después de esta conversación la persona no logra determinar si la otra parte es una máquina, entonces, en caso de ser una máquina, se le puede considerar inteligente [52].

Una máquina que sea capaz de pasar el Test de Turing debe contar con las siguientes características [52]:

- Reconocimiento del lenguaje natural:
Es una rama de la IA que, esencialmente, pretende extraer información de manera subjetiva y lograr la comprensión del mensaje que se intenta expresar, esta información puede presentarse en diferentes formas como texto o audio, a la cual se le aplican algoritmos para su interpretación [53].
- Razonamiento:
Éste es la vinculación entre juicios para obtener nueva información, es decir, extraer conclusiones y aprender en base a estos resultados [54].
- Aprendizaje:
Baldwin Hergenhahn define el aprendizaje como “un cambio re-

lativamente permanente en la conducta o en su potencialidad que se produce a partir de la experiencia y que no puede ser atribuido a un estado temporal inducido por enfermedad, fatiga o drogas” [55].

- Representación de conocimiento:
Esta característica manifiesta la comunicación del conocimiento de una manera que facilite su entendimiento como si se expresara en lenguaje natural, para que a su vez ésta contribuya a la creación de nuevo conocimiento. [56].

Con el paso del tiempo, se han considerado 2 características más para esta prueba, las cuales son [57]:

- Visión artificial:
Esta característica utiliza una cámara con la cual se harán capturas, posteriormente se procesarán dichas imágenes, donde el procesamiento descompone la imagen, con el fin de estimar diferencias entre píxeles, para determinar la presencia de alguna característica en particular [57] [58].
- Robótica:
Esto con el fin de poder manipular objetos y que dicha máquina inteligente, tenga movilidad propia [57].

Al considerar estas características, el “*Turing Test*” es llamado “*Total Turing Test*” [57], sin embargo, no es estricta la aplicación de estas nuevas características para considerar a una máquina inteligente.

2.6.1. Algoritmos Predictivos

Los algoritmos predictivos, comprenden una variedad de técnicas que predicen resultados futuros, basados en datos históricos y actuales.

En la práctica, se puede aplicar a casi todas las disciplinas, éstos buscan descubrir patrones y capturar relaciones en los datos [59]. Dicho de otra manera, un algoritmo predictivo es un conjunto de heurísticas y cálculos, que se utilizan para crear un modelo a partir de datos; para esto, el algoritmo realiza la extracción de los patrones previamente especificados, estos parámetros son aplicados a todo el conjunto de datos en repetidas ocasiones, para determinar resultados óptimos [60]. Hay que considerar que los resultados de la predicción, pueden verse afectados por factores como la cantidad de datos históricos disponibles o la calidad de éstos; puede parecer que el análisis predictivo es igual a hacer un pronóstico, pero existe una gran diferencia, mientras que un pronóstico puede predecir cuántas ventas se tendrá en el siguiente mes, el análisis predictivo puede indicar qué cliente es el que tiene mayor probabilidad de consumir en el siguiente mes, este resultado puede representar un punto de inflexión, si se utiliza de manera adecuada [61].

La interacción entre los diferentes marcadores o variables a considerar, representa un cálculo muy complejo para la mente humana, que si bien, puede realizar dicho cálculo, es poco factible. Es gracias a la aplicación de este tipo de algoritmo que se puede combinar información de manera sistemática y producir resultados confiables [62].

Para realizar una aplicación adecuada de este tipo de algoritmo es necesario cumplir con varias fases, que son [62]:

- Adquisición de datos:
Como primer medida es necesario la adquisición de datos provenientes del mundo real, es decir, que estos datos no provengan de una fuente ficticia o incorrecta. Si esta fuente no contiene información relevante, el análisis final no tendrá veracidad.
- Preprocesamiento:
Es necesario realizar una revisión a la fuente de información

antes de ser sometida al algoritmo, pues ésta puede contener valores atípicos que deben ser eliminados.

- **Extracción de características relevantes:**
Ésta es una de las fases más importantes, en este punto se deben considerar todas las características relevantes que el algoritmo aplicará.
- **Selección de modelo/Entrenamiento:**
Es aquí donde se selecciona el algoritmo o algoritmos a utilizar, por otra parte, el algoritmo seleccionado debe ser entrenado con el conjunto de datos.
- **Evaluación:**
Es el último componente, en este punto es necesario realizar una validación externa de los resultados, es decir, comprobar la capacidad de predicción del algoritmo utilizando un nuevo dato de entrada, esto es extremadamente importante para garantizar que el algoritmo funcione para todos los datos de entrada y no sólo para los usados en el entrenamiento.

Actualmente, existen varios tipos de algoritmos para el reconocimiento de patrones y la predicción de datos o eventos, como lo son las redes neuronales, las máquinas de soporte vectorial, algoritmos de regresión lineal, redes bayesianas, entre otros.

Redes Neuronales

Una red neuronal es un procesador distribuido con capacidad de almacenar conocimiento en base a sus experiencias. Las redes neuronales se basan en la biología del cerebro humano, éstas imitan en cierto modo los procesos de aprendizaje de un cerebro humano principalmente en dos aspectos, el conocimiento es adquirido por la red a través de un proceso de aprendizaje, y las conexiones neuronales conocidas como ponderaciones sinápticas se utilizan para almacenar

el conocimiento [63].

La red consiste en una capa de entrada, una capa de procesamiento y una capa de salida, como se muestra en la figura 2.3.

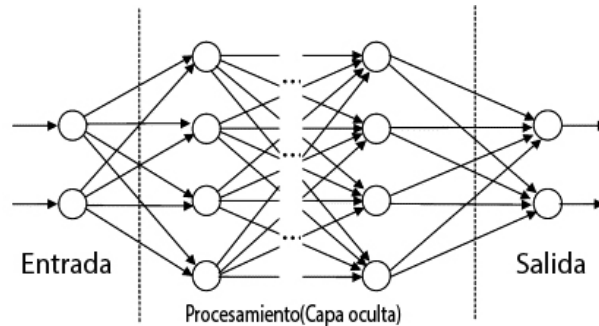


Figura 2.3: Arquitectura de una red neuronal.

Cada neurona está conectada a otras neuronas de una capa anterior, a través de conexiones sinápticas adaptables; las redes pueden ser utilizadas como método alternativo en el análisis y predicción de datos, sin embargo, si se solicitan resultados a una red que no a sido entrenada, los resultados son aleatorios y no tienen significado [63].

Según Eleyan & Demirel definen a las redes neuronales como “Herramientas poderosas que pueden ser entrenadas para realizar una compleja y diversas funciones en aplicaciones de visión artificial, tales como pre procesamiento (extracción de límites, restauración de imágenes, filtrado de imágenes), extracción de características (extracción de características de dominio transformadas), memoria asociativa (almacenamiento y recuperación de información), y reconocimiento de patrones” [64].

Máquina de soporte vectorial

Este algoritmo se basa en el aprendizaje en muestras para la realización de clasificadores y regresores, fundamentalmente, este algoritmo fue desarrollado para la resolución de problemas de clasificación binaria, sin embargo, actualmente es utilizado para resolver otro tipo de problemas como regresión, agrupamiento, y multi clasificación; debido a los sólidos fundamentos que integran este algoritmo. Mientras la mayoría de los algoritmos de aprendizaje se centran en obtener el menor porcentaje de error, este algoritmo se centra en la disminución del riesgo estructural, el aprendizaje estadístico también es incorporado y en conjunto con el clásico hiperplano óptimo de separación, confiere a la máquina de soporte una capacidad de generación superior a otros métodos de aprendizaje [65] [66]. Para el correcto aprendizaje se suele separar en dos grupos de datos, donde el primer grupo prestará la función de entrenamiento, y el segundo grupo de datos funcionará para realizar las pruebas para corroborar el patrón generado.

En la figura 2.4 se muestra la estructura de una máquina de soporte vectorial, donde se muestra el hiperplano de separación, el margen existente entre el hiperplano y el denominado vector de soporte, puede considerarse como el dato más cercano al hiperplano.

Éste algoritmo tiene una ventaja sobre la mayoría de los algoritmos de aprendizaje, pues presenta un bajo costo computacional, puede trabajar con un gran número de variables, sin importar el tipo de éstas, y la interpretación de sus resultados suele ser fácil de comprender [67].

Regresión Lineal

El modelo analiza la relación que existe entre la variable de respuesta y el conjunto de variables predictoras, su principal objetivo es la selección de parámetros que presenten de manera conjunta el menor porcentaje de error; gran parte del trabajo que realiza este algoritmo se centra en la minimización del error, así como asegurarse que las variables de predicción mantengan relación con la variable de respuesta [61].

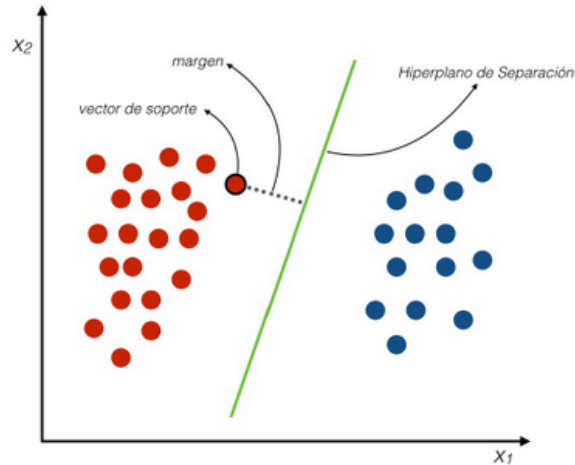


Figura 2.4: Datos linealmente separables.

La aplicación de este algoritmo tiene como principal objetivo [68]:

- Identificar:
Variables que puedan representar un papel de predictoras en relación con la variable de respuesta.
- Describir:
La relación que presentan las variables predictoras en relación con la variable de respuesta.
- Predecir:
La variable que se presentará como respuesta en base a las variables predictoras.

Redes Bayesianas

Las redes bayesianas utilizan métodos probabilísticos que permiten realizar predicciones, basado en los datos previamente recopilados,

e inferir la predicción a una variable desconocida o también conocido como un nuevo dato de entrada, entre las características que posee este algoritmo se puede destacar el conocimiento que logran adquirir sobre relaciones de dependencia y causalidad, esto les permite utilizar bases de datos incompletas.

Este algoritmo puede realizar trabajos de clasificación; bajo ciertos parámetros puede realizar predicciones, siempre y cuando la predicción sólo sea para una variable, mientras que el resto de variables son los datos propios de clasificación.

El proceso de aprendizaje de las redes bayesianas contempla dos etapas [69]:

- Aprendizaje estructurado:
consiste en obtener la estructura, es decir las relaciones de dependencia e independencia entre las variables.
- Aprendizaje paramétrico:
tiene como finalidad obtener las probabilidades y condicionales requeridas.

Reconocimiento de Patrones

Se define como reconocimiento de patrones a la clasificación de señales particulares de objetos físicos o abstractos. Para el reconocimiento de patrones se sigue un proceso de reconocimiento, de extracción de características y clasificación [70], se estima que el 80 % del trabajo que realiza una red neuronal se basa en el reconocimiento de patrones[71].

Para que el reconocimiento de patrones se considere efectivo, al menos debe cumplir con 3 requisitos:

- Una alta tasa de comprensión.
- Buena capacidad discriminatoria.
- Robustez frente a la interferencia.

El manejar los datos de una manera adecuada, facilita el proceso de toma de decisión y mejora la tasa de generalización. Sin embargo, se requiere tener un dominio profundo de la naturaleza del problema, lo cual suele presentar dificultades [71].

2.7. Trabajos Relacionados

Finalmente, esta sección describirá brevemente algunas de las aplicaciones que han utilizado algoritmos de IA en el ámbito de las adicciones.

2.7.1. AutoOPU

Según el Instituto Nacional sobre el Abuso de Drogas, en 2014, 18893 estadounidenses murieron a causa de analgésicos opioides y 10574 personas murieron por sobredosis de heroína, lo cual refleja un aumento significativo desde 2001. Esta es una enfermedad crónica que requiere tratamiento, el cual tiene un alto porcentaje de recaída y abandono ??.

La herramienta AutoOPU se centra en detectar automáticamente a los usuarios de opioides que utilizan la red social de Twitter, basado en el algoritmo de IA llamado “multi-kernel learning”, en el cual cada uno de sus núcleos, está ponderado automáticamente por el algoritmo de aprendizaje para hacer predicciones ??.

2.7.2. Aprendizaje automático como un complemento de la toma de decisiones clínicas en el tratamiento de dependencia del alcohol

Esta prueba se realizó con el fin de validar el uso de la IA como complemento en el ámbito médico, con el fin de ayudar a combatir el alcoholismo.

Dicha prueba presenta la evaluación de 73 sujetos dependientes al alcohol, los cuales durante 12 semanas participaron en un programa de terapia cognoscitiva (CBT), la cual ayuda a individuos a modificar su comportamiento del pensamiento de una manera que les ayude para llevar una vida que sea tan normal como sea posible, por otra parte, 66 sujetos más aplicaron un programa idéntico, sin embargo, a estos últimos no se les brindó el agente preventivo de recaídas Acamprosate, éste agente es un fármaco que está indicado en la terapia de mantenimiento de abstinencia en pacientes dependientes del alcohol previamente desintoxicados.

Aplicando 2 algoritmos de inteligencia artificial, como lo son los clasificadores bayesianos y los árboles de decisión, y utilizando como base los atributos de zona demográfica, uso del alcohol, nivel de dependencia, nivel de ansia y calidad de vida, estos algoritmos obtuvieron resultados que muestran un 73 % de acierto en la predicción de ambos algoritmos aplicados a la muestra que utilizó el agente Acamprosate, mientras que en el conjunto que no utilizó el agente se presentaron resultados del 77 % en referencia a los árboles de decisión y un 69 % en el conjunto bayesiano [72].

2.7.3. NeuCube

Esta metodología propuesta en 2015, presentó su prueba de funcionalidad en diferentes grupos de sujetos, entre los cuales se encontraban sujetos con adicción a los opiáceos; paciente que reciben tratamiento

de mantenimiento con metadona y un grupo de usuarios con drogas o control saludable, a los cuales se les realizó un electroencefalograma (EEG), esto con el fin de analizar los cambios funcionales en la actividad cerebral.

El método propuesto clasificó con mayor precisión los datos del EEG que los métodos tradicionales de IA y estadística, este método se puede usar para predecir la respuesta al tratamiento y el efecto del fármaco relacionado con la dosis. Pero lo que es más importante, el método puede ser utilizado para comparar las actividades cerebrales funcionales de diferentes sujetos y los cambios de estas actividades como resultado del tratamiento [73].

2.7.4. Trigr Health

Esta herramienta surge en base a las necesidades que muchas personas dependientes del alcohol y los opiáceos presentan, como lo es el no contar con alguien disponible en todo momento para expresarse. Este rol, normalmente es presentado por amigos y en ocasiones familiares, sin embargo, este método presenta fallas, pues es complicado brindar ayuda remota a una persona que ya se encuentra ansiosa por el uso de alguna sustancia; al utilizar esta aplicación como una herramienta de ayuda, se puede tener, en base al comportamiento, una predicción con días de anticipación con más del 92 % de precisión [74].

2.7.5. Predicción del uso de sustancias en los medios sociales

Actualmente, las personas gastan una cantidad significativa de tiempo en las redes sociales como facebook, interactuando con amigos y familiares, compartiendo un gran número de momentos o reaccionando a contenido de interés, es por ello que se utilizó esta fuente de información para realizar este estudio, y demostrar cómo las técnicas de aprendizaje automático y la técnica de minería de texto, pueden

ser utilizadas para construir sistemas efectivos de detección de sustancias, enfocados en medios sociales.

Utilizando métodos de reconocimiento de patrones en una gran cantidad de información no supervisada, en combinación con métodos basados en redes neuronales para el aprendizaje autodidacta, se logró obtener un 86 % en la predicción del uso de tabaco, 81 % en el uso de alcohol y 84 % en el uso de drogas [75].

Capítulo 3

Análisis y diseño del sistema

3.1. Introducción

Para la realización del sistema propuesto en la presente tesis, se utiliza como metodología el Proceso Unificado de Desarrollo de Software (PUDS), en el cual se emplea un modelo de arquitectura 4+1 vistas [76].

3.2. Análisis del sistema

3.2.1. Diagramas de contexto: Nivel 0

La figura 3.1 muestra el diagrama de contexto de nivel 0 que se busca modelar.

En este diagrama se presentan las interacciones que realiza un sistema con su entorno, los cuales pueden ser otros sistemas, internos o externos al mismo. El diagrama de contexto 0 tiene un alto nivel dentro de los diagramas de flujo de datos y debe de presentar únicamente un proceso. El proceso en el cual está enfocado el diagrama se representa mediante un rectángulo con las esquinas redondeadas, a este proceso se le asigna el número cero, mientras que las entidades



Figura 3.1: Diagrama de contexto: nivel 0.

externas son representadas por rectángulos; el flujo de datos se representa mediante flechas que señalan la dirección en la que fluyen los datos.

En la figura 3.1 se muestra el diagrama de contexto del sistema de detección y reconocimiento de patrones en bases de datos de adicciones. En este diagrama, el proceso principal es la aplicación del algoritmo, mientras que las entidades externas son las entradas de datos y la salida de estos datos para su posterior análisis.

3.2.2. Diagrama de contexto: Nivel 1

El diagrama de contexto de nivel 1, tienen la función de ampliar el panorama que el diagrama anterior presenta.

En la figura 3.2 se muestra el diagrama de nivel 1, el cual se basa en el diagrama analizado con anterioridad, este diagrama conserva las entidades externas, pero añade los procesos de lectura, selección y resultados, de igual manera se considera la base de datos. El flujo de datos inicia al ingresar un nuevo tamizaje, una vez cargado, se almacena la información en base de datos, a continuación se realiza la integración de datos, y a éstos se les aplica el vector de características. En el siguiente proceso se aplica el algoritmo predictivo, al terminar este proceso se realizan dos actividades, almacenar la información en la base de datos y generar resultados para su análisis e interpretación.

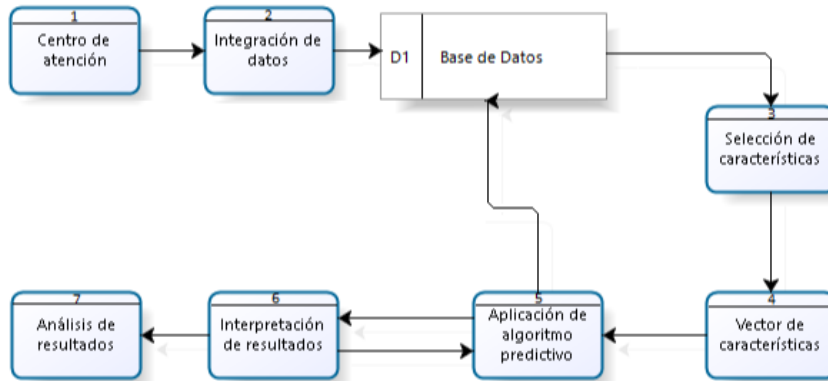


Figura 3.2: Diagrama de contexto: nivel 1.

3.2.3. Casos de uso

El lenguaje unificado de modelado (UML) brinda la oportunidad de modelar el sistema utilizando casos de uso, en los que se muestran las diferentes vistas del sistema, desde la perspectiva de los diferentes usuarios. En estos diagramas se describen las acciones que realiza un sistema, sin mencionar de qué manera lo hace.

En la figura 3.3 se presenta el diagrama de casos de uso que satisface los requerimientos del sistema a desarrollar como parte de la presente investigación. Se especifican un total de seis casos de uso principales (Login, Tamizajes, Catálogos, Preprocesamiento, Implementación de algoritmo predictivo y Análisis de resultados) con sus respectivos casos de uso secundarios.

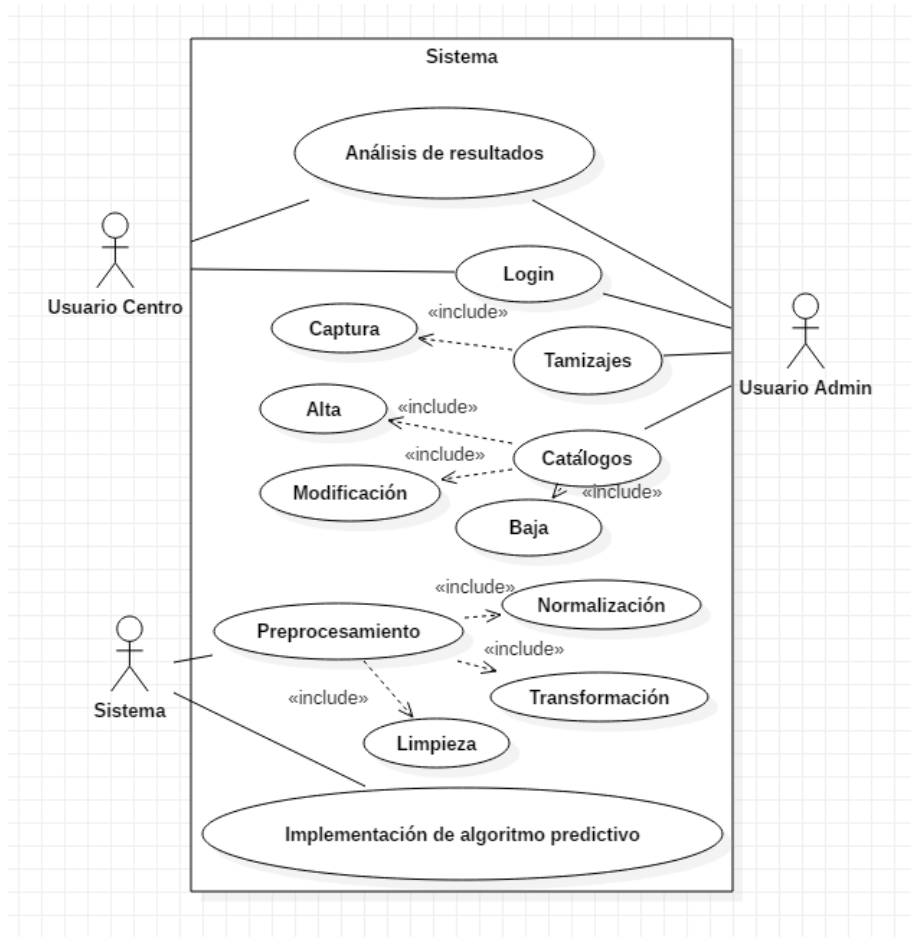


Figura 3.3: Diagrama de casos de uso del sistema.

A continuación se desarrollan los seis casos de uso previamente presentados: Login, Tamizajes, Catálogos, Preprocesamiento, Implementación de algoritmo y Análisis de resultados. Estos casos de uso se representarán en forma de diagramas de secuencia, en los cuales se describe el cómo y en qué orden, un grupo de objetos funcionan en conjunto.

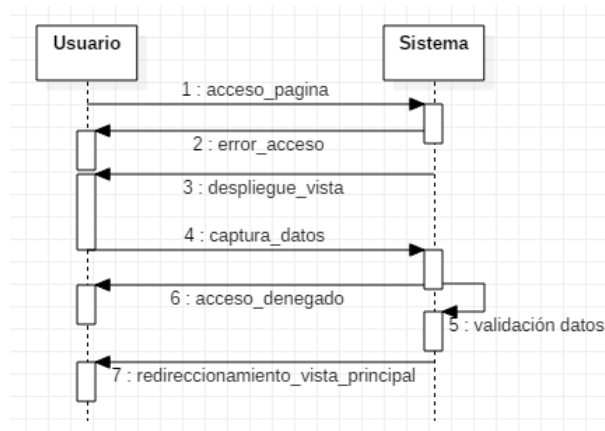


Figura 3.4: Diagrama de secuencia de caso de uso Login.

En la figura 3.4 se describe el flujo para acceder al sistema de acuerdo al perfil de usuario. En este diagrama participan dos de los usuarios mencionados con anterioridad, el usuario del centro y el usuario administrador.

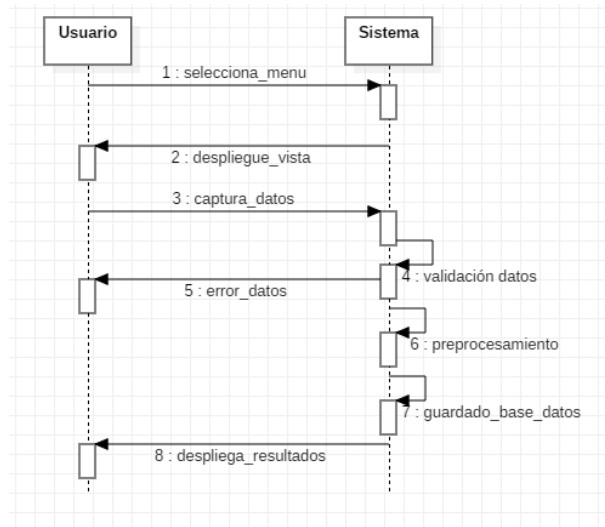


Figura 3.5: Diagrama de secuencia de caso de uso Tamizaje.

En la figura 3.5 se presenta el flujo que se seguirá al capturar un tamizaje. La participación de usuario se limita sólo al usuario administrador.

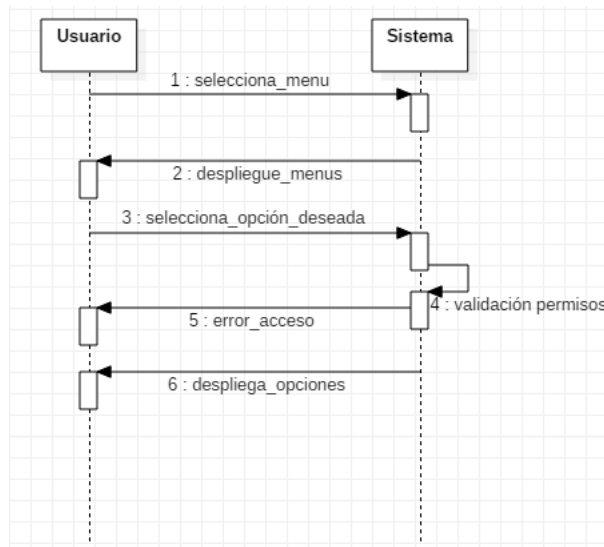


Figura 3.6: Diagrama de secuencia de caso de uso Catálogos.

La figura 3.6 muestra los procesos para trabajar con los catálogos. Particularmente, este caso de uso cuenta con procesos internos que se describen en las figuras 3.7 y 3.8. Estas figuras representan el alta de un nuevo dato en los catálogos y la actualización de alguno de sus registros. La participación de usuario se limita solo al usuario administrador.

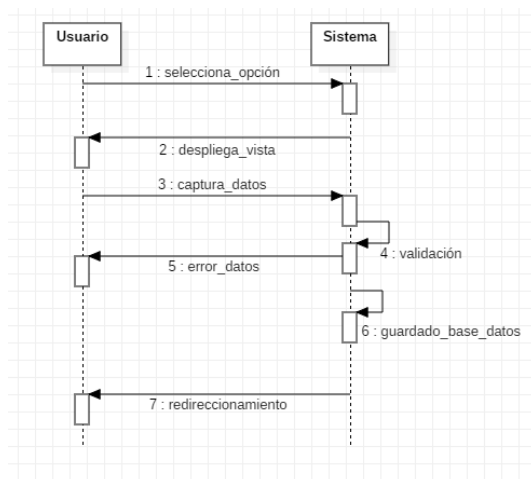


Figura 3.7: Diagrama de secuencia de caso de uso Catálogo extendido agregar.

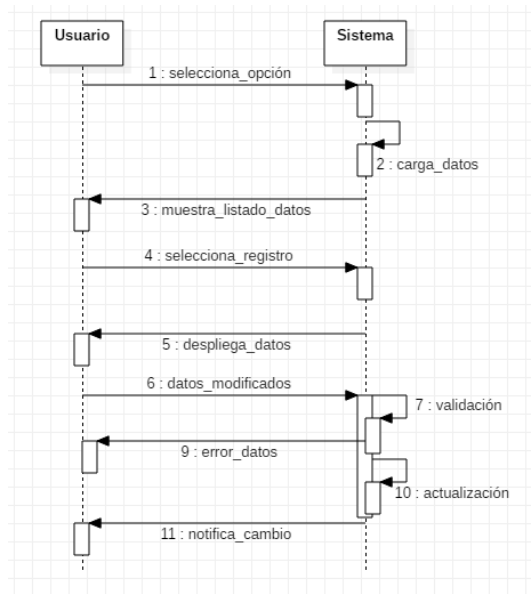


Figura 3.8: Diagrama de secuencia de caso de uso Catálogo extendido editar y/o eliminar.

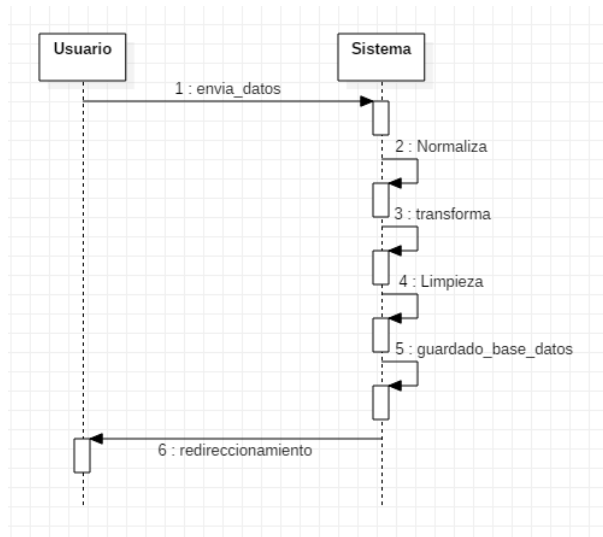


Figura 3.9: Diagrama de secuencia de caso de uso Preprocesamiento.

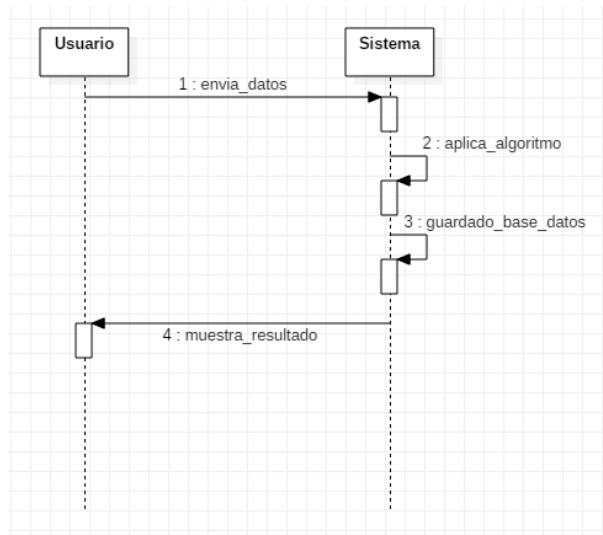


Figura 3.10: Diagrama de secuencia de caso de uso Implementación de algoritmo.

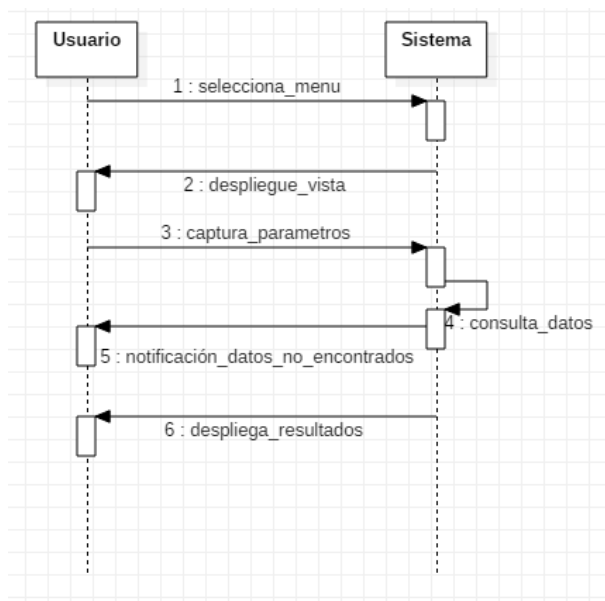
Secuencia Análisis de resultados:

Figura 3.11: Diagrama de secuencia de caso de uso Análisis de resultados.

En la figura 3.11 se muestra el flujo que los datos siguen para presentar los resultados obtenidos. Los actores en este caso de uso son el usuario del centro y el usuario administrador.

3.3. Diseño del Sistema

3.3.1. Diagrama de Actividades

En el siguiente apartado se presentan los diagramas de actividades, que explican el flujo de trabajo realizado por el usuario, estos flujos de trabajo pueden ser secuenciales, concurrentes o de ambos tipos y son representados por una flecha dirigida. Este tipo de diagrama utiliza distintos símbolos para representar las fases de las actividades. El ini-

cio y fin de una actividad están representados por un círculo de color verde y rojo respectivamente. Una actividad se representa mediante un rectángulo redondeado, este rectángulo contiene una descripción general de la actividad. El flujo entre actividades es controlado mediante decisiones, las cuales están representadas por un rombo y flechas, que indican las decisiones a seguir. También se utilizan carriles, éstos son utilizados para definir actividades en las que interactúan más de un subsistema o tipo de usuario [77].

Este diagrama representa cada caso de uso principal indicados en la figura 3.3

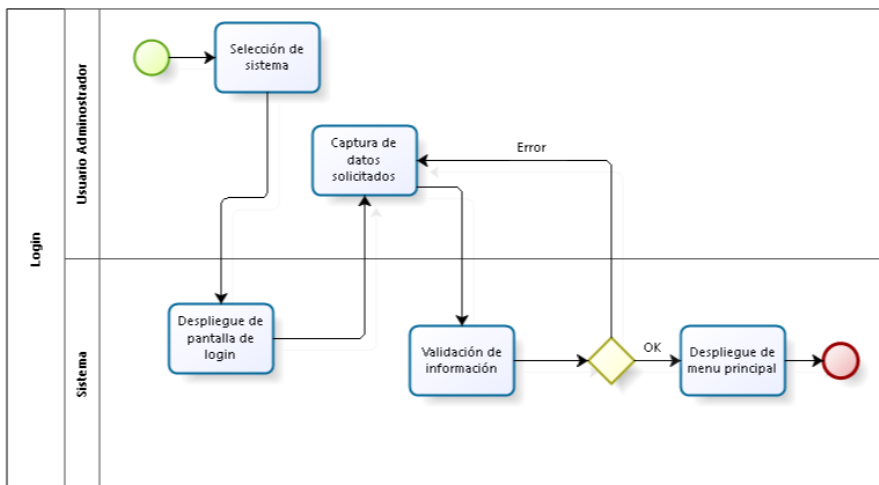


Figura 3.12: Diagrama de actividad para Caso de uso Login.

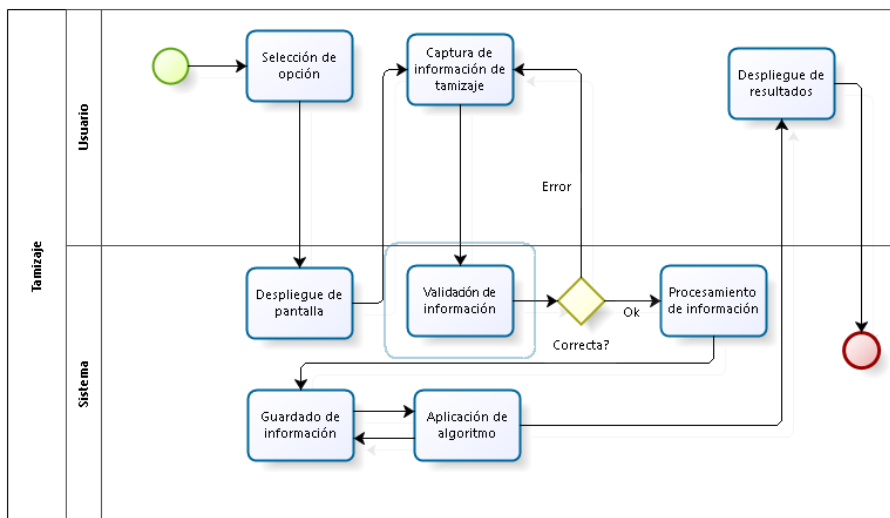


Figura 3.13: Diagrama de actividad para Caso de uso Tamizaje.

La figura 3.13 muestra el flujo que realiza el sistema, el flujo inicia al seleccionar la opción de captura de tamizaje, de manera seguida se valida la información capturada, si esta información es incorrecta, el usuario es reenviado a la pantalla de captura, en caso contrario, si la información es correcta se procede al guardado y aplicación del algoritmo, y finalmente, se despliega el resultado en pantalla.

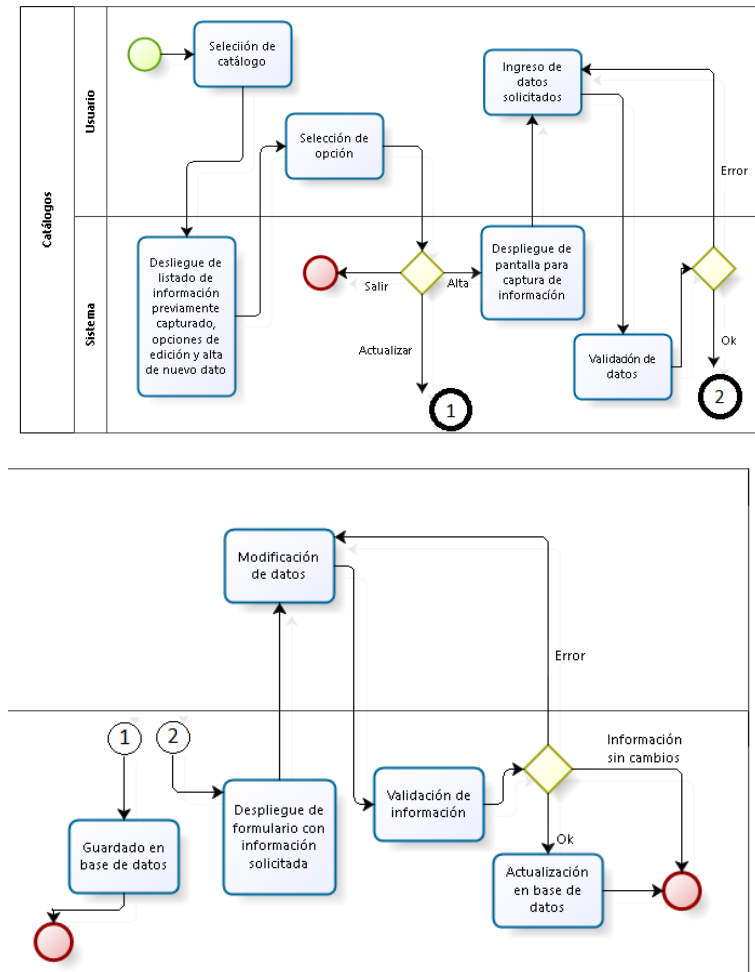


Figura 3.14: Diagrama de actividad para Caso de uso Catálogos.

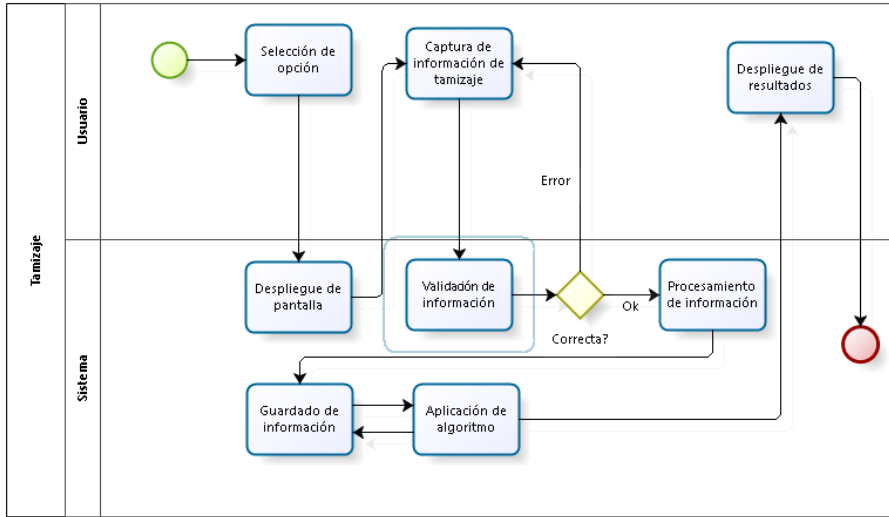


Figura 3.15: Diagrama de actividad para Caso de uso Análisis de resultados

3.3.2. Diagrama de Clases

Los diagramas de clases se utilizan para representar las características de un sistema, éstos tienen una gran utilidad, ya que trazan claramente la estructura de un sistema. Este tipo de diagrama se enfoca en presentar de manera clara las clases, métodos, atributos y la relación existente entre ellos. En este diagrama las clases están representadas por un rectángulo que se divide en tres filas. La fila superior contiene el nombre de la clase, la fila que se encuentra en el centro contiene todos los atributos de la clase, en éstos se puede especificar qué tipo de dato es y finalmente, la fila inferior contiene todos los métodos de la clase. En el diagrama de clases también se especifica el tipo de visibilidad que tendrán tanto los atributos como los métodos, representados por un símbolo de + para los públicos, un - para los privados y un # para los protegidos [78].

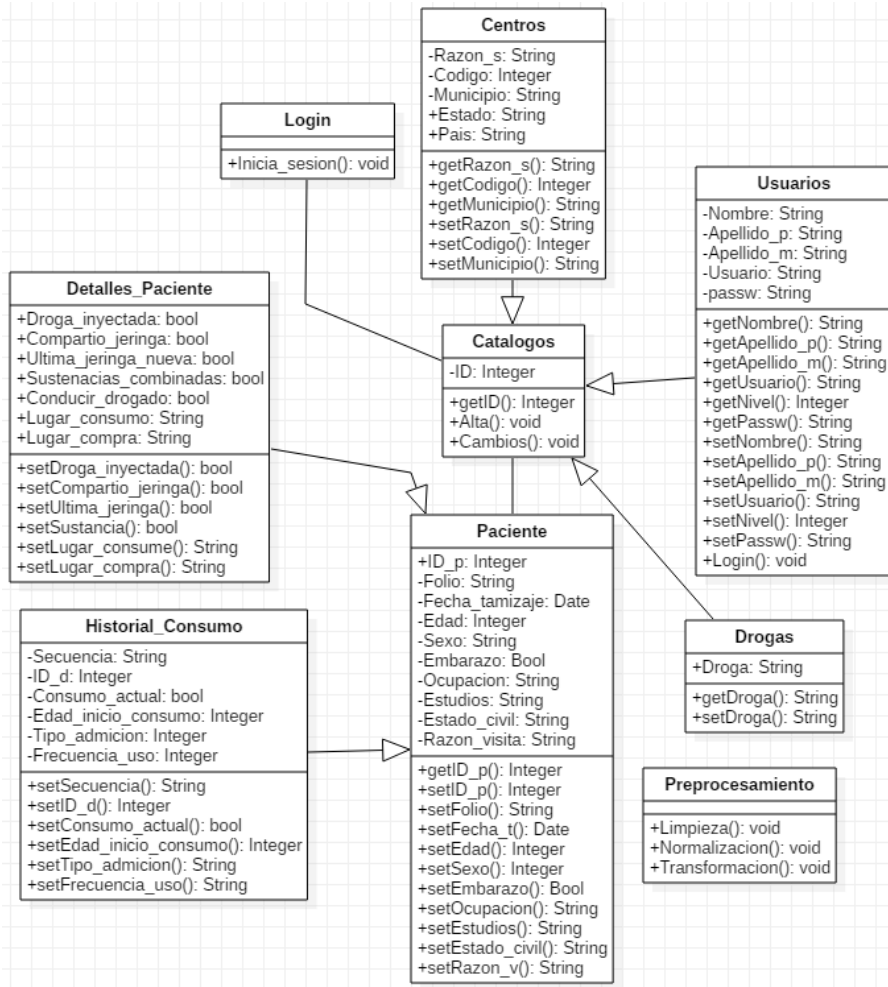


Figura 3.16: Diagrama de Clases

3.3.3. Diagrama de Componentes y Despliegue

El diagrama de componentes y despliegue es un diagrama que se utiliza para modelar el hardware utilizado en las implementaciones del sistema y la relación existente entre sus componentes. Además,

también es posible representar componentes físicos, bases de datos o módulos del sistema. El principal símbolo utilizado en este diagrama es el Nodo. Un nodo está representado por una caja o cubo y simboliza un recurso físico que contendrá al software. El software está representado dentro de un nodo mediante un cuadrado con el nombre del archivo y su extensión, ya sean una librería (ddl), un archivo ejecutable (exe), entre otros [78].

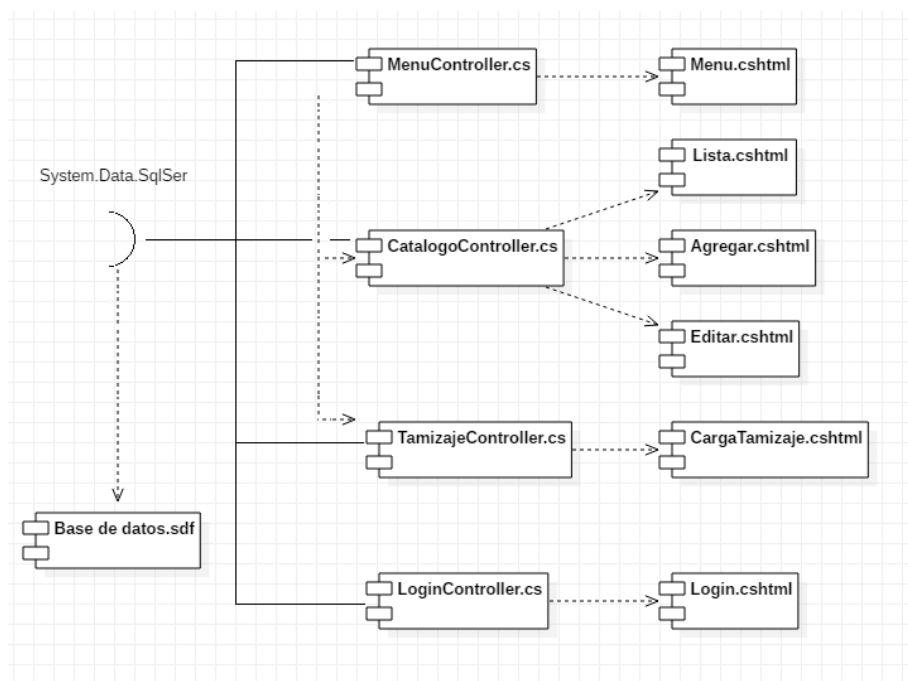


Figura 3.17: Diagrama de Componentes y despliegue.

3.4. Diagrama de base de datos

En esta sección se presenta el diagrama de base de datos que se utilizará para el guardado de los datos que genere el sistema. En este apartado se presentará el diagrama perteneciente a la base de datos.

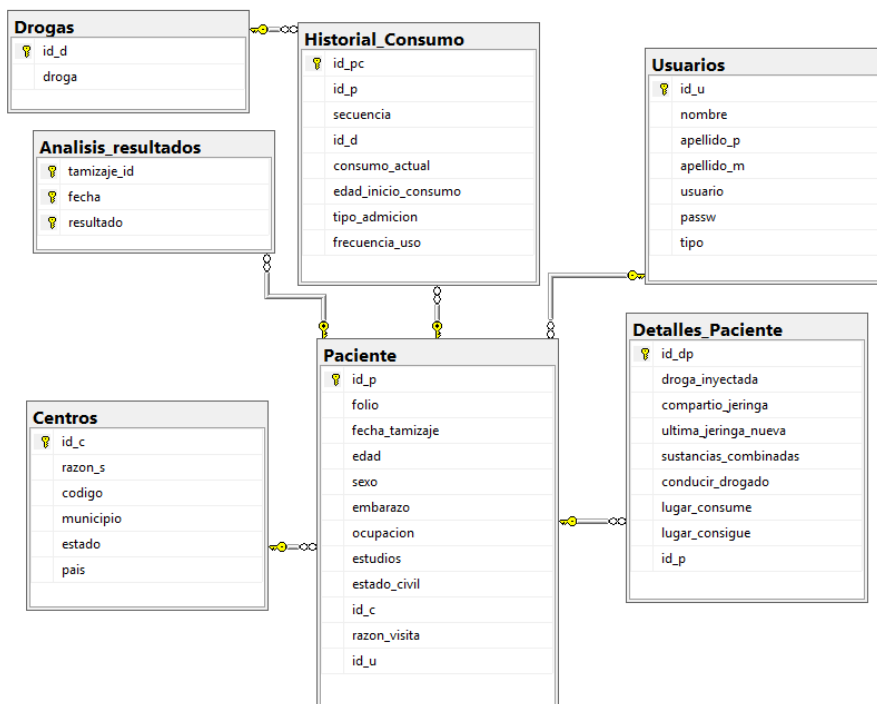


Figura 3.18: Diagrama de la Base de Datos.

En la figura 3.18 se muestra el diagrama de base de datos que el sistema utiliza, el cual contiene seis tablas; la tabla de Paciente guarda la mayor parte de la información referente al paciente, la tabla Historial consumo guarda la información de todo el consumo que el paciente ha descrito en el tamizaje, la tabla de Detalles_Paciente guarda los registros de un área específica del tamizaje, la tabla de Usuarios almacena a todos los usuarios que interactuarán con el sistema, la tabla Centros contendrá información de los diferentes centros en el estado que brindan atención, la tabla Drogas almacenará las diferentes drogas que son requeridas por el sistema y finalmente la tabla de Análisis_resultados mantendrá el registro de todos los tamizajes todos los tamizajes procesados por el algoritmo inteligente.

3.5. Arquitectura propuesta del sistema

A continuación se muestra la arquitectura propuesta para el sistema que describe los procesos que componen el sistema. La arquitectura presentada considera tres módulos: Módulo de lectura, Módulo de procesamiento y finalmente, Módulo de resultados.

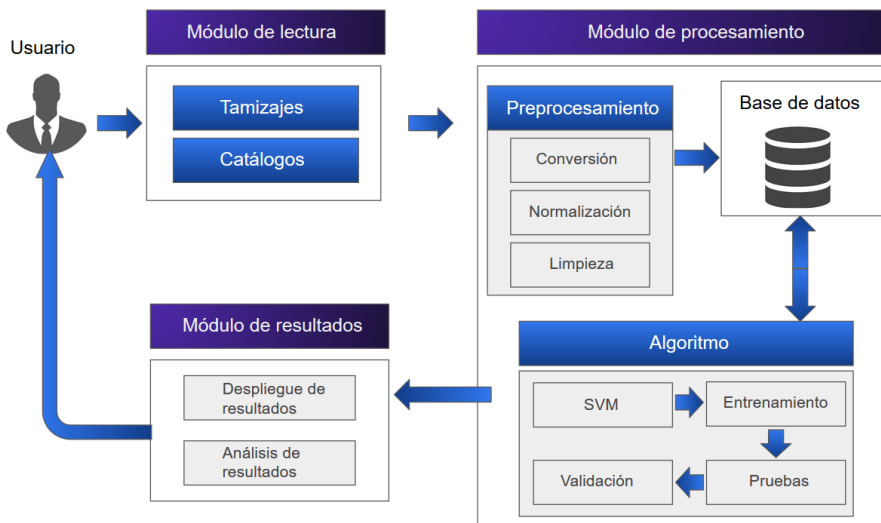


Figura 3.19: Diagrama de Arquitectura propuesta del sistema.

Módulo de lectura: Este módulo muestra la alimentación de información al sistema por parte del usuario, éste tendrá la posibilidad de interactuar con los subprocesos de catálogos y tamizajes.

Módulo de procesamiento: Este módulo realiza tareas diversas como son: preprocesamiento de los datos, la aplicación de el algoritmo de inteligencia artificial y la conexión y manejo de la base de datos.

Módulo de resultados: Este módulo se encarga de mostrar los resultados en pantalla, así como la retro-alimentación al Módulo de procesamiento.

Capítulo 4

Implementación

En este capítulo se presenta la implementación de las distintas etapas y pruebas realizadas para el desarrollo y validación del algoritmo propuesto, el cual está enfocado en la detección y reconocimiento de patrones en bases de datos de adicciones.

Se describe a detalle la recolección e integración de los datos, la selección de las características a utilizar, los diferentes *“datasets”* creados para las pruebas realizadas y hasta las distintas validaciones de los resultados obtenidos.

El entorno donde fueron realizadas las pruebas considera una computadora *“desktop”* con un procesador I7-4770, 16 gigabits de memoria ram ddr3 y una unidad de disco duro de estado sólido, en la cual se utilizó el entorno de anaconda con *“Python”* versión 3.

4.1. Recolección de los datos

En cuanto al conjunto de datos disponibles para la experimentación y validación, se tienen cinco formatos con extensión CSV (comma-separated values), que contienen la información recaudada en los años

2014, 2015, 2016, 2017, 2018 y 2019, por los diferentes centros de salud y manejada por el Ing. Aarón Rigoberto León Ruiz, así también se obtuvo la información referente al catálogo de sustancias psicoactivas.

Dichos formatos cuentan con la información referente al paciente, a sus antecedentes con respecto a las sustancias consumidas y el estado actual en el que se encuentra; esto con el fin de ayudar a conocer el tipo de tratamiento o seguimiento que se le dará al paciente. Estos tamizajes suelen ser individuales y bajo entrevistas clínicas [79].

Los tamizajes utilizados en la recolección de los datos, están divididos en dos formatos, el primero es considerado como un formato antiguo, ya es aquel que fue utilizado en los años 2014 a 2016; el segundo cuenta con el formato actual o actualizado, utilizado en los años 2017 a 2019.

A continuación se muestran los diferentes formatos previamente mencionados.

En la figura 4.1 se muestra la cara frontal del formato de tamizaje antiguo el cual corresponde a los años 2014 a 2016.

En la figura 4.2 se muestra la cara trasera la cual también es parte del tamizaje de los años 2014-2016.

En la figura 4.3 se contempla la cara principal del segundo formato de tamizaje referente a los años 2017-2019 el cual sustituyó al formato anterior.

En la figura 4.4 se muestra el complemento del formato de tamizaje actual, si bien estos formatos mantienen características similares, en su forma actualizada se presentan 42 nuevos atributos, gracias a esta adición, el nuevo formato de tamizaje cuenta con 80 atributos, mientras que en su formato anterior solo contaba con 48 de éstos.

Ambos tamizajes cuentan con una estructura similar, sin embargo, se presentan algunas diferencias, como se muestra en la figura 4.1; en el primer apartado se solicitan datos de identificación, en este apartado se captura si el paciente ha buscado con anterioridad ayuda en algún centro, en caso de que la respuesta sea positiva, también se le pregunta en cuántas ocasiones, mientras que en el formato actual que se presenta en la figura 4.3, se solicitan además, datos socio demográficos, específicamente, su ocupación actual, nivel de escolaridad y su estado civil, mientras que en el formato 2014-2016 el apartado socio demográfico se presenta como el segundo. Como tercer apartado, ambos formatos comparten el patrón de consumo, en este apartado se recolecta la información referente a las drogas consumidas, desde la droga de inicio hasta la décima droga consumida, en caso de existir el consumo de varias de éstas, también se captura la droga de impacto, una droga de impacto es la detonante en la solicitud de ayuda. Como complemento de esta información, también se captura la droga del listado consumida en el último año, la edad de uso por primera vez, la forma de administración, el consumo actual y cuántas veces se ha consumido.

El siguiente apartado se conoce como conductas de riesgo, este apartado está enfocado en conocer si el usuario tiene la posibilidad de estar infectado con una enfermedad de transmisión hemática, específicamente si se ha compartido alguna jeringa al momento de administrar alguna sustancia. De igual manera se realizan preguntas sobre la disponibilidad de la sustancia o droga, en este apartado se realizan un par de preguntas con la finalidad de conocer el lugar en el cual el paciente suele consumir la sustancia psicoactiva y sobre el lugar donde la consigue.

El formato actualizado, presenta un apartado nuevo en comparación con el formato 2014-2016, en éste se presenta el apartado de dependencia a la droga de impacto, este apartado tiene la finalidad de conocer el patrón de dificultades que se ha presentado con respecto

a la droga de impacto, la tolerancia que se ha desarrollado, el tipo de consumo hacia ésta, los síntomas de la abstinencia al disminuir o interrumpir su consumo y, finalmente, el uso persistente de ésta.

Si bien en ambos formatos se incluye información que podría ser considerada como sensible, ésta no es solicitada al paciente para mantener la confidencialidad de sus datos.

4.2. Análisis de atributos

Con el objetivo de encontrar las características con mejor efectividad para el procesamiento, se realizó un estudio de ambos tamizajes, para conocer el dominio de las características presentadas en cada uno de sus formatos, lo cual permite reducir la cantidad de información que será procesada y analizada.

Se analizaron y compararon los dos formatos de tamizajes existentes, como se muestra en la figura 4.2:

En el análisis de los atributos realizado, se tomaron en cuenta los atributos que se presentaban en ambos formatos, como se muestra en la figura 4.5, esto con la intención de generar “*datasets*” con una mayor cantidad de datos. Como parte de el análisis se tomaron como base los atributos que los expertos en el tema quienes pertenecen a los centros de salud recomendaron, y que son: edad, sexo, estado civil y ocupación principal.

Posteriormente se detectaron otros atributos en común como lo son el nivel de estudios, el centro, la droga de inicio, el lugar de consumo y finalmente, el lugar de compra.

De igual manera, como parte del análisis de los atributos se consideraron la segunda, tercera, cuarta, quinta, sexta, séptima, octava,

novena y décima droga, pero estos fueron descartados ya que la gran mayoría de los datos presentaban un gran número de nulos, vacíos y valores indefinidos. Otros atributos que fueron descartados son los que se encuentran en el apartado de conductas de riesgo, ya que éstos, si bien proporcionan información relevante para conocer si el paciente presenta problemas referentes a alguna enfermedad hepática, no proporcionan un gran valor para nuestro vector de características.

Una vez que se analizaron los atributos de ambos tamizajes, se realizó el cargado de los archivos previamente mencionados en el manejador de base de datos “*SQL Server*”, mediante el asistente de lectura de archivos, se crearon las tablas correspondientes a cada archivo y sus atributos, una vez cargados en el manejador de la base de datos, se procedió al procesamiento de éstos, con el fin de agrupar todos los datos en una misma tupla y se mantuviera la estructura y tipo de los atributos.

Finalmente, se crearon los diferentes “*datasets*”, donde se realizó la validación de los atributos para que no contaran con datos que pudieran crear anomalías.

4.3. Experimentación

Al haber realizado un análisis a profundidad de los formatos de tamizajes y la información recolectada, se procedió a realizar experimentaciones con diferentes atributos.

El procedimiento utilizado para la lectura y extracción de la información de los “*datasets*” creados se muestra en la figura 4.6

```
1  Algoritmo Lectura de dataset
2  -----
3  Entradas
4  url: ruta donde se encuentra el archivo de lectura.
5  names: nombres de asignación para cada una de las características del vector.
6  dataset: lectura del archivo y asignación de nombre a las columnas.
7
8  Salidas:
9  Variable con la información del archivo.
10
11 Procedimiento:
12 Especificación del archivo a leer.
13 Asignación del nombre de las cabeceras del vector.
14 Leer información del archivo y asignacion de la cabeceras (url, names).
```

Figura 4.6: Algoritmo de lectura de “*dataset*”

El algoritmo que se muestra en la figura 4.6 describe la lectura de el archivo especificado, el cual cuenta con una extensión .CSV, este formato contiene la información concentrada de los tamizajes, se le asigna el nombre a cada una de las columnas generadas y finalmente, se realiza una asignación de esta información a una variable.

En la tabla 4.1 se muestran los datos obtenidos en la lectura realizada.

Cuadro 4.1: Distribución, agrupado por clases.

Droga	Valor	Casos	Droga	Valor	Casos
Cerveza	10101	1061	Psilocibina	40901	1
Destilados	10103	167	Diacetilmorfina	50101	2016
Pulque	10104	2	Opio	60101	24
Vino	10105	11	Morfina	60201	1
Alcohol Puro	10106	23	Metadona	60204	13
Otros	10107	2	MST	60301	1
Marihuana	20101	1213	Durogesic	60305	9
Pegamentos	30101	6	Fentanest	60306	2
Pegamento	30103	21	Cocaína	70101	643
Adhesivo	30106	14	Loracepam	80101	3
Cola	30109	1	Rivotril	80118	209
Resina	30111	1	Tafil	80120	16
Sprays	30201	1	Ritalin	100101	23
Tinner	30403	26	Dexedrina	100102	1
Afrinex Infantil	40101	1	Asenlix	100203	3
Antiflu-des	40102	5	Redotex N F	100226	1
Atarax	40104	1	Belladona	110101	2
Cheracol	40107	1	Tabaco	120101	44
Loratadina	40108	1	Cocaína	150101	71
Cortafin	40114	1	Metanfetamina	160101	14981
LSD	40201	6	Rohypnol	170101	2
DMT	40401	1	Cocaína	180101	232
Polvo de Angel	40701	34	—	—	—

La tabla 4.1 muestra los resultados obtenidos al ejecutar el algoritmo. Cabe señalar que debido a la legibilidad de la tabla fue separada en dos columnas, donde se muestra el listado de las drogas agrupado y contabilizado por clase.

El algoritmo utiliza solo las columnas de valor y casos, sin embargo, se agregó la columna droga, con la intención de dar a conocer cual

es la descripción del valor. Como parte de los resultados obtenidos, en la primer columna se muestra la droga Cerveza, con el código identificador 10101 y presenta 1061 casos de los 20898, también se puede observar una alza en casos como la Metanfetamina que presenta 14981 casos, la Diacetilmorfina con 2016 casos y la Marihuana con 1213 casos, esto por mencionar las drogas que, en la lectura de los archivos realizada, han orillado a más pacientes a solicitar ayuda.

La figura 4.7 muestra la configuración de las métricas para el funcionamiento del algoritmo, dicho funcionamiento se explica a continuación.

```
1 Algoritmo configuración de metricas.
2 -----
3 Entradas
4 array: volcado de información del dataset.
5 x: selección de las columnas a utilizar para el entrenamiento.
6 y: selección de la columna que se tomará como clase de salida.
7 validation_size: tamaño de la muestra que se tomará para realizar la validación.
8 X_train: separación de la variable x para entrenamiento.
9 X_validation: separación de la variable x para la validación.
10 Y_train: separación de la variable y para entrenamiento.
11 Y_validation: separación de la variable y para la validación.
12 scoring: se especifica qué tipo de puntuación se busca.
13
14 Salidas:
15 Asignación de la configuración para los valores a procesar.
16
17 Procedimiento:
18 Se realiza la descarga de información del dataset en una nueva variable.
19 Se especifica qué columnas de la matriz se utilizarán como entrada.
20 Se especifica qué columna de la matriz se utilizará como salida.
21 Se especifica el tamaño de la muestra que se tomará para la validación.
22 Se divide el dataset en los parámetros previamente asignados tanto de entrenamiento
23 como de validación.
24 Se especifica la puntuación.
```

Figura 4.7: Algoritmo de configuración de métricas.

Como parte del proceso realizado por el algoritmo, se configuran las métricas del conjuntos de entrenamiento y validación, como se muestra en la figura 4.7, se definen los atributos a considerar para el entrenamiento como X y el atributo que se definirá como salida

es definido como Y , así también se define el tamaño del conjunto de validación, es decir, qué porcentaje del conjunto de datos será utilizado para realizar las validaciones del funcionamiento, una vez definido este conjunto, se define de manera automática el conjunto de entrenamiento.

```
1 Algoritmo ejecución de los algoritmos de IA.
2 -----
3 Entradas
4 models: se define un vector vacío para los modelos.
5 results: se define un vector vacío para los resultados.
6 names: se define un vector vacío para la asignación de los nombres.
7
8 Salidas:
9 Resultados de los modelos evaluados.
10
11 Procedimiento:
12 Se realiza la Asignación del algoritmo de regresión lineal.
13 Se realiza la Asignación del algoritmo de vecino más cercano.
14 Se realiza la Asignación del algoritmo de árboles de decisión.
15 Se realiza la Asignación del algoritmo de redes bayesianas.
16 Se realiza la Asignación del algoritmo de máquina de soporte vectorial.
17 Se realiza un ciclo para la evaluación de cada uno de los modelos asignados.
18 Se evalúa cada modelo.
19 Se asigna el resultado obtenido a la variable results.
20 Se asigna el nombre del algoritmo ejecutado.
```

Figura 4.8: Algoritmo de asignación y de ejecución.

Con el fin de utilizar la mejor opción con respecto a la problemática presentada, se realizaron pruebas a diferentes algoritmos de IA. En la figura 4.8 se presenta el algoritmo de ejecución, donde también se muestran los diferentes algoritmos de IA que se compararon, entre ellos se encuentran, regresión logística (LR), vecino más cercano (KNN), árboles de decisión (CART), redes bayesianas (NB) y máquina de soporte vectorial (SVM).

Como parte del algoritmo de ejecución que se describe en esta figura, también se encuentra la evaluación que se realiza para cada uno de los algoritmos, a continuación se muestran los resultados obtenidos

de una de las ejecuciones realizadas.

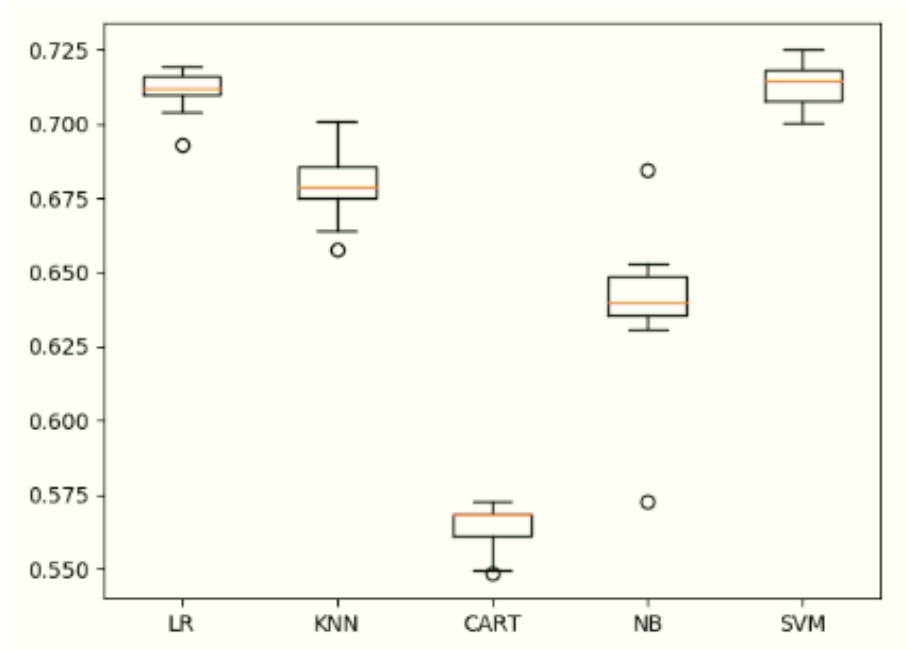


Figura 4.9: Gráfica de resultados en comparativa de los algoritmos.

La figura 4.9 muestra el valor máximo obtenido (línea superior de la caja) y mínimo (línea inferior), la media la cual se muestra con una línea naranja al interior del rectángulo, y finalmente el rango de valores en los cuales se mantuvo el entrenamiento.

Cuadro 4.2: Resultados de comparativo de los algoritmos.

Algoritmo	% Precisión
Regresión logística	0.7109
Vecino más cercano	0.6790
Árboles de decisión	0.5639
Redes bayesianas	0.6387
Máquina de soporte	0.7201

El cuadro 4.2 muestra los resultados de los algoritmos, donde la regresión lineal y la máquina de soporte vectorial presentaron los mejores resultados con un 71 y 72% respectivamente, mientras que los árboles de decisión arrojaron el porcentaje más bajo con 56%.

Sin embargo antes de obtener estos resultados se realizaron varias pruebas o corridas de los diferentes *“datasets”*, las cuales se describen a continuación.

4.3.1. Primera experimentación con datos de 2017

En esta experimentación se utilizó un *“dataset”* el cual contiene 6330 registros del año 2017. En esta primera experimentación se utilizaron sólo los atributos de edad, sexo, estado civil y ocupación, que fueron recomendados por los expertos. El algoritmo utilizado en cada una de las experimentaciones es similar, sin embargo, es necesario realizar modificaciones en las métricas que definen las columnas de entrada y la columna de salida, así como el archivo a utilizar, dicho cambio se presenta en la siguiente figura.

La figura 4.10 muestra el código utilizado para realizar las experimentaciones de esta investigación, desde la lectura del archivo hasta obtener los resultados, este código es descrito por las figuras 4.6, 4.7 y 4.8.

Como parte de las líneas específicas utilizadas en la primera expe-

rimentación, línea 3, se especifica el nombre del archivo a leer, en la figura 4.10 se muestra el archivo con nombre `dataset(5)_17` con extensión `csv`, el cual hace referencia a un **“dataset”** con cinco atributos y los datos del año 2017. Otra línea que se requiere es la número 4, en esta línea se especifican los nombre de las columnas. Finalmente, las líneas 11 y 12 especifica el número de columnas que se utilizarán como atributos de entrada y el atributo de salida, respectivamente.

El resultado obtenido al ejecutar la corrida con el **“dataset”** previamente mencionado arrojó un 64 % de precisión.

4.3.2. Segunda experimentación con datos de 2017-2018

La segunda experimentación se realizó con un **“dataset”**, que contiene 13696 registros de los años 2017 y 2018 y un **“dataset”** con 36638, el cual contiene los datos de los años de 2014 a 2018, éstos cuentan con los atributos recomendados por los expertos.

Para esta experimentación fue necesario modificar la línea número 4 que se presenta en la figura 4.10, esto para cambiar el **“dataset”** a leer, el resto de las líneas se mantiene ya que se tuvo un incremento de datos y no de características.

Al realizar la corrida de estos **“datasets”**, se obtuvieron 69 % y 64 % respectivamente, cabe resaltar que con el **“dataset”** con los datos de los años 2014 a 2018, el cual contaba con más de 30,000 registros no presentó un incremento en el porcentaje obtenido en comparación con la primera experimentación.

4.3.3. Tercera experimentación con datos de 2014-2018

En esta experimentación se tomó como base los “*datasets*” previamente mencionados, es decir, cuenta con las 4 características recomendadas por los expertos, pero fue agregado el atributo de Estudio. Estos “*datasets*” contaban con 13696 registros en sus años 2017 y 2018, y 36513 en los años 2014 a 2018.

La ejecución de esta experimentación requirió modificaciones en las líneas de lectura del archivo, específicamente en el nombre de las columnas, las columnas donde se definen las columnas de entrada y la columna de salida, es decir, se modificaron las líneas 4,5,10 y 11 como se muestra en la siguiente figura.

```
3 url = "dataset(6)_all_19.csv"
4 names = ['Edad', 'Sexo', 'Estado Civil', 'Ocupacion', 'Estudios', 'Droga']
10 X = array[:,0:5]
11 Y = array[:,5]
```

Figura 4.11: Líneas de código actualizadas.

La figura 4.11 muestra los cambios requeridos para la ejecución de esta experimentación, en ella se muestran las líneas mencionadas con anterioridad, como lo son, la línea número 3, la cual indica que se utilizó un archivo con nombre dataset(6)_all_19 con extensión csv, este nombre indica que este archivo contiene todos los datos recolectados desde el año 2014 al año 2018 y cuenta con 5 atributos. Otra de las líneas que fue necesario actualizar, es la línea número 4, en donde se agregó la columna Estudios. Finalmente, se encuentran las líneas 10 y 11, en donde se define la cantidad de columnas a utilizar como datos de entrada y de salida respectivamente.

Los resultados obtenidos al realizar esta experimentación fueron de 70% con el “*dataset*” de los años 2017 y 2018, mientras que el

“*dataset*” referente a los cinco años no presentó un incremento y se mantuvo en un 64 %.

4.3.4. Cuarta experimentación con datos de 2014-2018

En la cuarta experimentación se tomó el Centro como un nuevo atributo para ambos “*dataset*”, sin embargo, como se mencionó con anterioridad, fue necesaria la edición de las líneas 3, 4, 10 y 11 para la lectura y ejecución del algoritmo de manera similar a como se muestra en la figura 4.11, en este caso, el archivo utilizado lleva el nombre de `dataset(7)_17_18` y `dataset(7)_all`, estos archivos hacen referencia a la información de los años 2017 y 2018, mientras que el segundo contiene la información de los años 2014 a 2018.

Al realizar la corrida de este algoritmo con los “*datasets*” correspondientes, se obtuvieron 71 % y 66 %, lo cual presentó un incremento de 1 % y 2 % respectivamente, con respecto a la tercera experimentación.

4.3.5. Quinta experimentación con datos de 2014-2018

En esta experimentación se utilizó la característica de droga de inicio o droga inicial para robustecer los “*datasets*” creados para las pruebas, estos contenían un total de siete atributos, que son: edad, sexo, estado civil, ocupación principal, estudios, centro y finalmente, la droga de inicio, estos contaban con 13689 y 36650 registros respectivamente.

La preparación requerida para realizar la corrida de esta experimentación utilizó las líneas 3, 4, 10 y 11 como se muestra en la figura 4.11, estas líneas fueron adaptadas a las necesidades presentadas por

esta experimentación, las cuales contemplaron, la agregación del nuevo atributo, el cambio del archivo de lectura y finalmente la configuración de los atributos de entrada y salida.

Los resultados arrojados por esta experimentación no presentaron cambios con respecto a la experimentación número cuatro, es decir, los resultados obtenidos por esta experimentación fueron de 71 % y 66 %, respectivamente.

4.3.6. Sexta experimentación con datos de 2014-2018

Para la sexta experimentación se realizaron tres escenarios, uno donde los “*datasets*” contemplan el lugar de compra como nuevo atributo, estos “*datasets*” contienen los siete atributos utilizados en la experimentación anterior y el agregado recientemente, es decir, cada uno de éstos contienen ocho atributos y sus archivos contienen 13671 registros para los años 2017 a 2018 y 36611 para los datos del 2014 a 2018. El segundo escenario, sustituye el atributo de droga de inicio agregado en la experimentación anterior por el atributo de lugar de compra, es por ello que los “*datasets*” utilizados en este escenario contienen ocho atributos y los archivos presentan 13689 y 36651 registros de los años 2017 a 2018 y del 2014 a 2018.

Finalmente el tercer escenario contempla los atributos de edad, sexo, estado civil, ocupación, estudios, droga de inicio y lugar de compra, en esta experimentación fue sustituido el atributo de centro por código de centro y al igual que el primer escenario contiene 13671 registros para los años 2017 a 2018 y 36611 para aquel del 2014 a 2018.

Todos los escenarios requirieron su configuración única para realizar la corrida del algoritmo, sin embargo, la configuración requerida es similar a la ya presentada en las experimentaciones anteriores, donde se especifican las métricas de lectura del archivo, el los nombres del

vector y las variables de entrada y salida.

Los resultados arrojados por los tres escenarios fueron del 71 % para los años 2017 a 2018 y 66 % al contemplar los 5 años.

4.3.7. Séptima experimentación con datos de 2014-2018

La séptima experimentación contempla dos escenarios, el primero contempla los atributos de edad, sexo, estado civil, ocupación, estudios, lugar de compra y lugar de consumo mientras que el segundo contempla edad, sexo, estado civil, ocupación, estudios y lugar de consumo, Los *“datasets”* utilizados en el primer escenario contienen 13601 registros para los años 2017 a 2018 y 36514 para los años 2014 a 2018.

Los resultados para esta experimentación en ambos escenarios, tuvieron los mismos valores, es decir, 71 % y 66 % respectivamente.

4.3.8. Octava experimentación con datos de 2014-2019

Para esta experimentación se obtuvo un año más de datos, los cuales representan al año 2019, este año adiciona 7103 nuevos datos. Para la realización de estos *“datasets”* se consideraron solo los cuatro atributos recomendados por los expertos, de manera similar a la experimentación número dos, pero enriquecidos con la información referente al año 2019, estos *“datasets”* contaban con 21003 registros para los años 2017 a 2019 y 41103 para aquel que contiene los años 2014 a 2019.

La configuración de las métricas fue similar a la de la segunda experimentación.

Los resultados en esta experimentación arrojaron 71 % para los tres últimos años y 65 % para aquel que contenía los seis años.

4.3.9. Novena experimentación con datos de 2014-2019

De manera similar a la experimentación número tres, en esta experimentación se utilizaron los cuatro atributos recomendados por los expertos y se sumó el atributo de centro, para continuar las pruebas discriminatorias. Esta experimentación al igual que la anterior cuenta con la adición de los 7103 datos del año 2019. Estos cuentan con 21003 registros en los años 2017 a 2019 y 41103 para los años 2014 a 2019.

Los resultados arrojados por estas experimentaciones fueron de 72 % y 66 %.

4.3.10. Validación del algoritmo utilizado

Debido a los resultados obtenidos, se optó por corroborar el correcto funcionamiento del algoritmo, para esto se realizaron pruebas utilizando un “*dataset*” de Iris setosa, éste es un “*dataset*” clásico en el ámbito de la inteligencia artificial.

Este “*dataset*” cuenta con 150 muestras y 5 atributos, los atributos que se contemplan son el largo y ancho del sépalo, el largo y ancho de los pétalos, estas medidas son presentadas en centímetros; finalmente, como último atributo se presenta la especie a la que pertenece la flor, la figura 4.12 presenta gráficamente las 3 especies y de qué manera se obtuvieron el resto de los atributos.



Figura 4.12: Iris Setosa.

El resultado obtenido en esta prueba para validar el correcto funcionamiento del algoritmo utilizado fue de 99 %, el cual fue el resultado esperado, ya que las diversas publicaciones presentan un valor similar. Con esto se demuestra la correcta aplicación del algoritmo. Sin embargo; como parte de las validaciones, se realizaron experimentaciones en diferentes software enfocados a IA como lo son Orange y WEKA.

En éstos se utilizaron diferentes *“dataset”* para las pruebas, en el software de Orange se utilizó el mismo *“dataset”* que se presentó en la cuarta experimentación, es decir, el archivo utilizado contaba con las características de edad, sexo, estado civil, ocupación, estudio y centro, mientras que en el software de WEKA, se utilizó un *“dataset”* las características previamente mencionadas, y fueron añadidas las características de droga de inicio, lugar de consumo y lugar de compra. Esto último, debido a que WEKA cuenta con una función que detecta cuáles son los atributos que mayor aportación tienen en cuanto al resultado y realiza su propia recomendación.

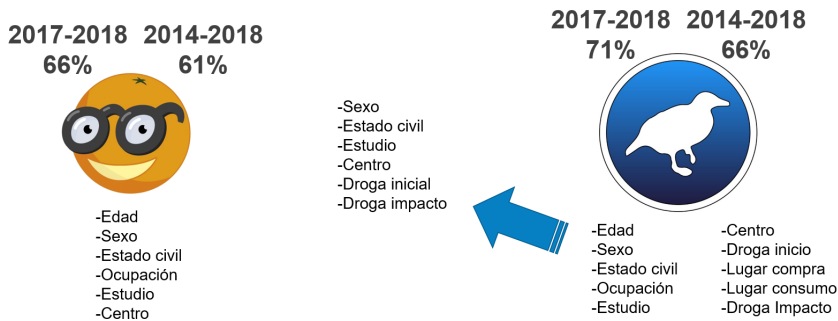


Figura 4.13: Experimentación con Orange y WEKA.

La figura 4.13 muestra los resultados obtenidos por ambos programas. El programa Orange arrojó resultados con un 5 % de decremento en ambos resultados, sin embargo, fue consistente en el resto de la pruebas decrementando el mismo valor. Los resultado del programa WEKA reflejan un resultado similar a las experimentaciones realizadas en el algoritmo desarrollado para la aplicación, como parte de los atributos seleccionados por el programa WEKA, se presentaron Sexo, Estado civil, Estudios, Centro y Droga de inicio, sin embargo, los resultados no reflejaron cambios al utilizar estos atributos.

Como parte de las pruebas realizadas se incluyen escenarios con los diferentes kernels que ofrece el algoritmo de SVM, los cuales son: Lineal, Polynomial, RBF , Sigmoid y finalmente auto, los cuales ajustan la línea de soporte para una mejor la separación de clases, así también, se utilizó una validación cruzada para la obtención del resultado.


```
1 Algoritmo ejecución con kernel auto.
2 -----
3 Entradas
4 tiempo_inicial: Asignación de la fecha y hora de inicio del proceso.
5 svc: Asignación de tipo de kernel.
6 tiempo_final: Asignación de la fecha y hora de la conclusión del proceso.
7 tiempo_ejecucion: Asignación de la diferencia entre los tiempos de inicio y fin del proceso.
8 cross_val_score_svc: Asignación del resultado obtenido.
9
10 Salidas:
11 Resultados de los modelos evaluados.
12
13 Procedimiento:
14 Se realiza la captura de la fecha hora desde el inicio del proceso.
15 Se realiza la Asignación del tipo de kernel(linear).
16 Se realiza el entrenamiento del algoritmo de SVM.
17 Se realiza la validación cruzada.
18 Se realiza la captura de la fecha hora al terminar el proceso.
19 Se realiza la compración entre la hora inicial y la hora final.
```

Figura 4.14: Código de ejecución de la validación cruzada.

La figura 4.14 muestra el algoritmo del código utilizado para la ejecución de la validación cruzada, en ésta, se muestra la inicialización de una variable para calcular el tiempo de ejecución que toma cada una de las corridas del algoritmo, de manera seguida se define el tipo de kernel a utilizar y se cargan los datos de entrenamiento, enseguida se ejecuta la validación cruzada, finalmente, se realiza el cálculo entre el tiempo de inicio de la operación y el tiempo final, para obtener el tiempo de ejecución.

4.4. Capturas del Sistema

En cuanto al sistema desarrollado como parte de la presente investigación, se contemplaron los módulos y funcionalidades previamente mencionados en el capítulo tres, a continuación se muestran algunas capturas de pantalla.

En la figura 4.15 se muestra la pantalla que se utiliza para la autenticación de usuarios, la cual tiene la función de validar las credenciales del usuarios que intenta acceder al sistema.

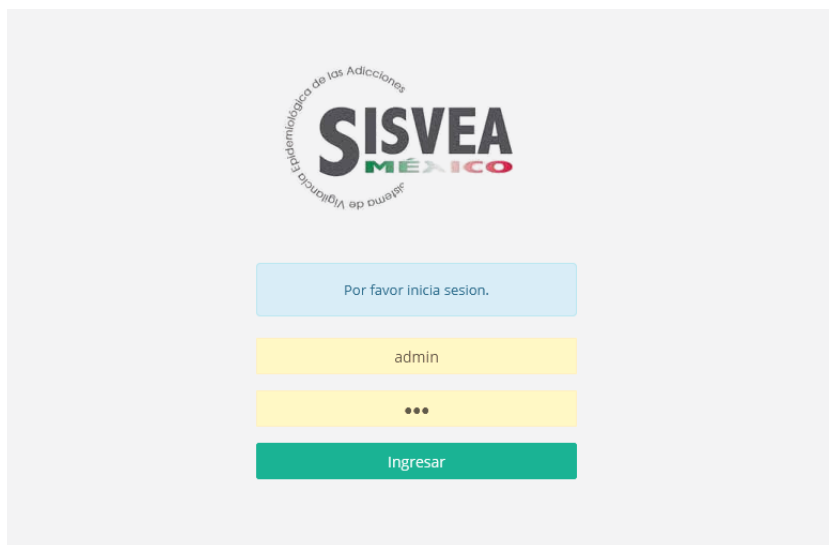


Figura 4.15: Captura de la pantalla de login.

Como muestra la figura 4.15 es necesario rellenar los campos de usuario y administrador, en caso de que el usuario no presente credenciales validas, es arrojado un mensaje de alerta donde se da el aviso de la denegación de acceso, por otra parte, si la información es correcta, el usuario es direccionado a la pantalla principal del sistema 4.16.

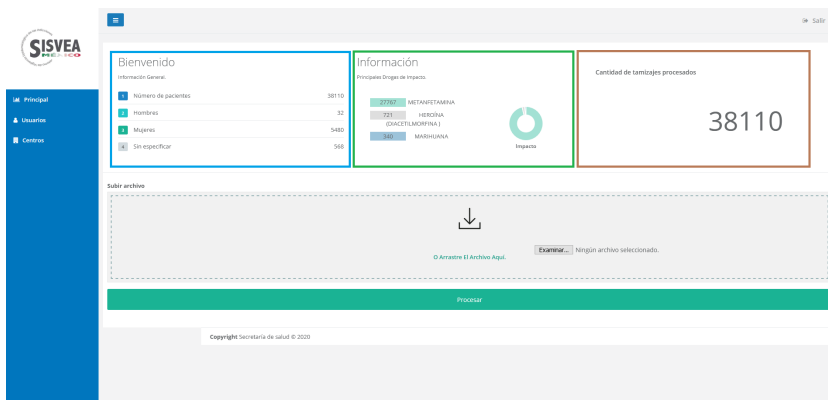


Figura 4.16: Captura de la pantalla principal.

La figura 4.16 muestra tres recuadros de diferentes colores, el primero cuadro con color azul muestra de manera general la información de los pacientes que se encuentran registrados en el sistema, estos son separados en base a su sexo. El siguiente recuadro con coloración verde muestra el TOP tres de las drogas de impacto presentadas en los tamizajes guardados en el sistema; finalmente se muestra de color café la cantidad de registros procesados por el algoritmo, es decir, aquellos a los que el sistema ya arrojó un reconocimiento.

Al realizar el procesamiento del tamizaje (formato excel), se procede a guardar la información en la base de datos, validando que el número de folio no se encuentre previamente registrado, al terminar este guardado, se inicia el algoritmo de IA el cual arroja los resultados por Folio y Predicción, a su vez, este guarda los resultados obtenidos.

Los módulos de centros y usuarios funcionan de manera similar, sin embargo la pantalla de centro tiene una mayor funcionalidad, es por ello que es utilizada para la descripción.

Total de Lecturas Realizadas

Mostrar 10 registros

Buscar

Agregar Centro

ID	RAZON	CODIGO	MUNICIPIO
104	dentro 1	5542	URBS
99	AMIGOS PARA SIEMPRE A.C. I.A.P.	5592	CAJEME
98	CPV A.C. (CLÍNICA PARA MENORES PALO VERDE AC)	5591	HERMOSILLO
97	ADICTOS UNIDOS EN RECUPERACION A.C.	5586	SAN LUIS RIO COLORADO
96	CPV SOLIDARIDAD	5585	HERMOSILLO
95	NAVICONON NAVIQUA	5584	NAVIQUA
94	IDAD QUATMAS	5583	QUATMAS
93	AMARE A.C. APSON	5582	AGUA PRIETA
92	NUOVA CREACION PENASCO HOMBRES	5581	PUERTO PENASCO
91	UCIDAD SONORA A.C.	5580	HERMOSILLO

Mostrando registros del 1 al 10 de un total de 100 registros

Anterior 1 2 3 4 5 10 Siguiente

Figura 4.17: Captura del módulo de centro.

La figura 4.17 muestra el listado de todos los centros que actualmente se encuentran registrados la base de datos del sistema, los cuales pueden ser filtrados por el número de registros a mostrar en pantalla, esta funcionalidad es remarcada con un cuadro azul, ya que los datos se encuentran paginados y se limita a mostrar diez registros por “*default*”, sin embargo, este valor puede ser modificado a veinticinco, cincuenta o cien, los datos pueden ser recorridos también por páginas utilizando la funcionalidad que se muestra con el recuadro amarillo, otra de las funcionalidad que presenta este tipo de módulos es el buscador, el cual busca por las coincidencias escritas en el, este se muestra en el recuadro naranja; finalmente se muestran los recuadros verdes y rojos, estos tienen la funcionalidad de agregar un nuevo centro o mostrar los datos para su edición respectivamente

Al utilizar la funcionalidad de agregado o edición, se despliega un modal en pantalla para agregar un nuevo centro o editar la información del centro seleccionado para su edición.

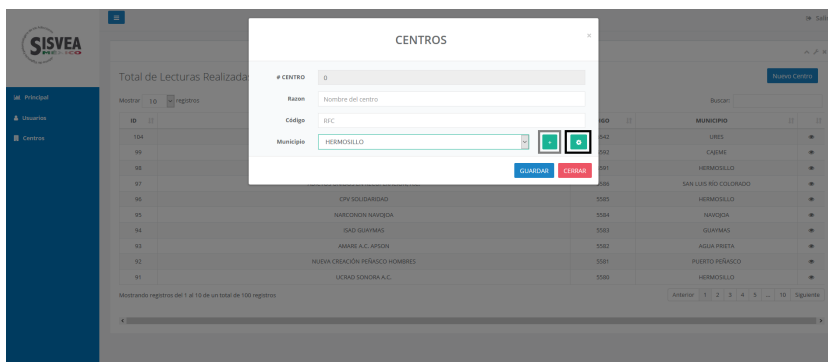


Figura 4.18: Captura de la pantalla para agregar un centro.

En la figura 4.18 se muestra el recuadro con los campos solicitados para agregar un nuevo centro, como parte de la funcionalidad de este modal, también es posible agregar o editar un municipio, esto sin necesidad de interrumpir el proceso de dado de alta o edición del centro, al agregar el nuevo municipio se recargada de manera automática el campo de selección de municipio, esto para que se despliegue en esta lista el nuevo registro.

MORELOS Servicios **SA**

SECRETARIA DE SALUD FOLIO [] [] [] []

SISTEMA DE VIGILANCIA EPIDEMIOLOGICA DE LAS ADICCIONES (SISVEA)
 CUESTIONARIO PARA CENTROS DE TRATAMIENTO Y REHABILITACION

Los datos del presente cuestionario son de carácter estrictamente confidencial y solo serán usados para contribuir a la planeación de programas de prevención y protección de la salud

FECHA 19 06 14
 Día Mes Año

I. DATOS DE IDENTIFICACIÓN: (Añote la respuesta sobre la línea sin ocupar recuadros).

Ciudad Obregon

Nombre del Centro de Tratamiento Amigos Para Siempre
 (Marque con "X" el cuadro con la respuesta correcta y/o escriba la respuesta en la línea).

1. ¿Há buscado anteriormente apoyo o atención en algún Centro de Tratamiento por el consumo de drogas?
 1 Sí 2 No Pase a la preg. 3

2. ¿Cuántas veces? _____

3. ¿En esta ocasión acude a este centro?
 1 Voluntariamente 2 Lo llevaron amigos o familiares 3 Por indicación en la escuela o el trabajo
 4 Por indicación del médico 5 Por indicación legal 6 Otros Especifique _____

II. CARACTERÍSTICAS SOCIODEMOGRÁFICAS DEL PACIENTE.
 (Añote la respuesta sobre la línea o marque con "X" la opción correspondiente a la situación del paciente)

Lugar de residencia actual:

4. Estado Sonora

5. Municipio Tajama

6. Localidad Obregon

7. Edad en años cumplidos 20

8. Sexo 1 Hombre 2 Mujer

9. Nivel escolar
 1 Sin estudios (No sabe leer ni escribir) (pase a la preg. 11)
 2 Sin estudios (Sabe leer y escribir) (pase a la preg. 11)
 3 Primaria 4 Estudios técnicos o comerciales
 4 Secundaria 5 Bachillerato o bachillerato técnico
 7 Estudios superiores 8 Estudios de posgrado

10. ¿Esos estudios?
 1 Los concluyó 2 No los concluyó 3 Los cursa actualmente

11. Estado civil
 1 Soltero (a) 2 Unión libre 3 Divorciado (a)
 2 Casado (a) 4 Separado (a) 6 Viudo (a)

12. Principal actividad
 1 Estudiante de tiempo completo 6 Desempleado (a) (busca empleo)
 2 Estudiante con actividad laboral 8 Desocupado (a) (No busca empleo)
 3 Con actividad laboral estable (más de 6 meses) 7 El Hogar
 4 Con actividad laboral inestable (menos de 6 meses) 8 Pensionado (a)

www.morelos.gob.mx

Figura 4.1: Cara frontal del formato de tamizaje de los años 2014-2016.

MORELOS Servicios de Salud

III. PATRÓN DE CONSUMO.
Escriba sobre las líneas su respuesta ó marque con "X" la opción correcta.
(Frecuencia anual detalladamente al orden en el que el paciente le consumió cada una de las sustancias)

TIPO DE SUSTANCIA (DROGA) ALCOHOL Y TABACO TAMBIEN SON DROGAS	ACTUALMENTE LA CONSUME (Último año)	EDAD EN QUE LA USO POR 1ª VEZ	FORMA DE ADMINISTRACIÓN PRINCIPAL	FRECUENCIA DE CONSUMO ACTUAL
1ª SUSTANCIA (DROGA) Tabaco	<input checked="" type="checkbox"/> SI <input type="checkbox"/> NO	14	1	2
2ª SUSTANCIA (DROGA) Alcohol	<input checked="" type="checkbox"/> SI <input type="checkbox"/> NO	14	4	5
3ª SUSTANCIA (DROGA) Marihuana	<input checked="" type="checkbox"/> SI <input type="checkbox"/> NO	18	1	1
4ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
5ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
6ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
7ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
8ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
9ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
10ª SUSTANCIA (DROGA)	<input type="checkbox"/> SI <input type="checkbox"/> NO			
SUSTANCIA (DROGA) DE IMPACTO (Es la que motivó al tratamiento) Marihuana	<input checked="" type="checkbox"/> SI <input type="checkbox"/> NO	18	1	1

FRECUENCIA

1 Diario: Más de 3 veces
2 Diario: De 2 a 3 veces
3 Diario: Una vez
4 De 2 a 3 veces por semana
5 Una vez a la semana
6 De 2 a 3 veces por mes
7 Una vez al mes
8 Menos de 1 al mes
9 Ya no la usa
10 Sólo la usó una vez

FORMAS DE ADMINISTRACIÓN

1 Fumada o inhalada
2 Aspirada (en mucosa nasal)
3 Inyectada
4 Ingesta
5 Tabaco (cigarrillos)
6 Aplicada en otras situaciones

24. ¿Alguna vez ha usado sustancias (drogas) inyectadas? SI NO Pase a la preg 27

25. ¿Alguna vez ha compartido una jeringa para inyectarse sustancias (drogas)? SI NO

26. La última vez que se inyectó sustancias (drogas), ¿La hizo con una jeringa que nadie más había usado antes de usted? SI NO

27. ¿Consumo sustancias (drogas) combinadas? SI NO Pase a la preg 29

28. ¿Cuáles principalmente? _____

29. No quiere que me diga el resultado, pero:
¿En los últimos 12 meses, le han hecho la prueba para detectar el virus del SIDA? SI NO Pase a la preg 31

30. ¿Sabe el resultado? SI NO

IV. DISPONIBILIDAD DE LA DROGA. (Marque con "X" la opción correcta)

31. ¿En que lugar consigue con más frecuencia la sustancia (droga) por la que acude a tratamiento?

1. Tiene en su propia casa

2. Se la llevan a su casa

3. En casa de un amigo o familiar

4. En lugares de distribución (la calle, parques, etc)

5. En discotecas/Antros/Bares

6. En fiestas o reuniones

7. En la escuela o el trabajo

32. ¿En que lugar acostumbra consumir, con más frecuencia la sustancia (droga) por la que acude a tratamiento?

1. En su propia casa

2. En casa de un amigo o familiar

3. En la vía pública (la calle, parques, etc)

4. En discotecas/Antros/Bares

5. En fiestas o reuniones


6. En la escuela o el trabajo

Nombre del paciente: Manriquez Pacheco Pedro Fecha de nacimiento: 10/07/1988
(apellido paterno) (apellido materno) nombre (s)

MUCHAS GRACIAS POR SU INFORMACIÓN

www.morelos.gob.mx

Figura 4.2: Cara trasera de formato del tamizaje de los años 2014-2016.



 Secretaría de Salud

 Dirección General de Epidemiología

 Sistema de Vigilancia Epidemiológica de las Adicciones

CENTROS DE TRATAMIENTO Y REHABILITACIÓN NO GUBERNAMENTALES

Fecha:

 Fecha de elaboración:

 Número de identificación:

DATOS SOCIODEMOGRÁFICOS

Fecha de nacimiento:

 Edad:

 Entidad de residencia:

 Municipio de residencia:

Anónimo: 1 Sí 2 No

 Sexo: 1 Masculino 2 Femenino

Nombre:

 Apellido paterno:

 Apellido materno:

 Nombre(s):

En relación con su actividad principal ¿cuál de las siguientes opciones se adecúa a usted?

<input type="checkbox"/> 1 Estudiante de tiempo completo	<input type="checkbox"/> 4 Tiene menos de 6 meses en su actual empleo	<input type="checkbox"/> 7 Hogar	<input type="checkbox"/> 10 Oba
<input type="checkbox"/> 2 Estudiante con actividad laboral	<input type="checkbox"/> 5 Desempleado (busca empleo)	<input type="checkbox"/> 8 Pensionado	<input type="checkbox"/> 88 No sabe
<input type="checkbox"/> 3 Tiene empleo estable de más de 6 meses	<input type="checkbox"/> 6 Desocupado (no busca empleo)	<input type="checkbox"/> 9 No aplica	<input type="checkbox"/> 99 No responde

En relación con su escolaridad ¿cuál es su máximo nivel de estudios?

<input type="checkbox"/> 1 Primaria	<input type="checkbox"/> 3 Escolaridad técnica	<input type="checkbox"/> 5 Licenciatura	<input type="checkbox"/> 7 No sabe leer ni escribir	<input type="checkbox"/> 9 No aplica	<input type="checkbox"/> 99 No responde
<input type="checkbox"/> 2 Secundaria	<input type="checkbox"/> 4 Bachillerato	<input type="checkbox"/> 6 Maestría / Doctorado / Especialidad	<input type="checkbox"/> 8 Sabe leer y escribir	<input type="checkbox"/> 88 No sabe	

En relación con su estado civil ¿cuál de las siguientes opciones se adecúa a usted?

<input type="checkbox"/> 1 Soltero(a)	<input type="checkbox"/> 2 Casado(a)	<input type="checkbox"/> 3 Unión libre	<input type="checkbox"/> 4 Separado(a)	<input type="checkbox"/> 5 Divorciado(a)	<input type="checkbox"/> 6 Viudo(a)	<input type="checkbox"/> 7 No aplica	<input type="checkbox"/> 88 No sabe	<input type="checkbox"/> 99 No responde
---------------------------------------	--------------------------------------	--	--	--	-------------------------------------	--------------------------------------	-------------------------------------	---

Lugar del Centro de Tratamiento:

Entidad:

 Jurisdicción:

 Municipio:

Nombre del Centro de Tratamiento:

 Código del Centro de Tratamiento:

ANTECEDENTES DE BÚSQUEDA DE APOYO

¿En este año, ha acudido a este tipo de Centro o algún otro? 1 Es la primera vez en el año 2 Ya había acudido en este año 88 No sabe 99 No responde

¿En toda su vida, incluyendo esta ocasión ¿cuántas veces ha estado en este tipo de Centros? Número de veces:

¿En esta ocasión acude a este Centro?:

<input type="checkbox"/> 1 Voluntariamente	<input type="checkbox"/> 4 Lo llevaron amigos	<input type="checkbox"/> 7 Oba
<input type="checkbox"/> 2 Por indicación médica no psiquiátrica	<input type="checkbox"/> 5 Por indicación legal	<input type="checkbox"/> 88 No sabe
<input type="checkbox"/> 3 Por indicación médica psiquiátrica	<input type="checkbox"/> 6 Por indicación de la escuela/trabajo	<input type="checkbox"/> 99 No responde

PATRÓN DE CONSUMO Recuerde anotar detalladamente el orden en que el usuario fue consumiendo cada una de las sustancias o drogas

EL ALCOHOL Y EL TABACO TAMBIÉN SON DROGAS Y PUEDEN APARECER COMO DROGA DE INICIO E IMPACTO

Tipo de droga	Nombre	Consumo en el último año		Edad de uso por la vez	* Forma de administración	** Consumo anual	Cuántas veces
Droga de INICIO		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Segunda droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Tercera droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Cuarta droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Quinta droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Sexta droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Séptima droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Octava droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Novena droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Décima droga		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Droga de IMPACTO		<input type="checkbox"/> Sí	<input type="checkbox"/> No	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Lugar de consumo de la DROGA DE IMPACTO * Forma de administración ** Consumo actual

Entidad: <input type="text"/>	<input type="checkbox"/> 1 Fumada	<input type="checkbox"/> 4 Inyectada	<input type="checkbox"/> 7 Ingerida	<input type="checkbox"/> 88 No sabe	<input type="checkbox"/> 1 Diario	<input type="checkbox"/> 3 Mensual
Municipio: <input type="text"/>	<input type="checkbox"/> 2 Inhalada	<input type="checkbox"/> 5 Topica	<input type="checkbox"/> 8 Masticada	<input type="checkbox"/> 99 No responde	<input type="checkbox"/> 2 Semanal	<input type="checkbox"/> 4 Anual
	<input type="checkbox"/> 3 Intenso (enfriada)	<input type="checkbox"/> 6 Otros mucosas	<input type="checkbox"/> 9 Oba			

Figura 4.3: Cara frontal del formato de tamizaje de los años 2017-2019. [79]

CONDUCTAS DE RIESGO

¿Alguna vez ha usado sustancias o drogas inyectadas? 1 Sí 2 No 88 No sabe 99 No responde

¿Alguna vez ha compartido una jeringa para inyectarse sustancias o drogas? 1 Sí 2 No 88 No sabe 99 No responde

La última vez que se inyectó sustancias o drogas, ¿lo hizo con una jeringa que nadie más había usado antes que usted? 1 Sí 2 No 88 No sabe 99 No responde

¿Consumo sustancias o drogas combinadas? 1 Sí 2 No

¿Cuáles principalmente? 1) _____ 2) _____

¿Alguna vez ha conducido un vehículo de motor bajo el influjo de alcohol u otras drogas? 1 Sí 2 No 88 No sabe 99 No responde

¿Padece alguna enfermedad? Cual(es): _____

DISPONIBILIDAD DE LA SUSTANCIA O DROGA

¿En qué lugar consigue con mayor frecuencia la sustancia o droga por la que acude a tratamiento?

1 En su propia casa	5 En bares, clubes	9 En tiendas cercanas a mi casa	88 No sabe
2 Se la llevan a su casa	6 En fiestas o reuniones	10 Parques	99 No responde
3 En casa de un amigo o familiar	7 En la escuela	11 Picaderos	
4 En la calle	8 En el trabajo	12 Otro _____	

¿En qué lugar acostumbra consumir con más frecuencia la sustancia o droga por la que acude a tratamiento?

1 Tiene en su casa	4 En bares, clubes	7 En el trabajo	88 No sabe
2 En casa de un amigo o familiar	5 En fiestas o reuniones	8 En picaderos	99 No responde
3 En la vía pública (calles, parques, etc.)	6 En la escuela	9 Otro _____	

DEPENDENCIA A LA DROGA DE IMPACTO

En los últimos 12 meses

La siguiente sección evalúa la dependencia a la droga de impacto, por tal motivo en los espacios en blanco, debe de referir la sustancia que el usuario haya indicado como **DROGA DE IMPACTO**.

¿Usó en más de cinco ocasiones _____ Droga de Impacto _____ para estimularse, relajarse, sentirse mejor o sentirse más activo o alerta? 1 Sí 0 No

SI LA RESPUESTA ES AFIRMATIVA CONTINÚE

Sobreactivo

Después de la primera respuesta afirmativa (SI), pase a la pregunta 2a

1a. ¿Se dio cuenta de que tenía que usar más cantidad que antes de _____ Droga de Impacto _____ para lograr el efecto deseado? 1 Sí 0 No

1b. ¿Notó que la misma cantidad de _____ Droga de Impacto _____ le hacía menos efecto que antes? 1 Sí 0 No

1c. ¿Alguna vez se dio cuenta de que necesitaba más cantidad de _____ Droga de Impacto _____ para lograr el mismo efecto? 1 Sí 0 No

Compulsivo

Después de la primera respuesta afirmativa (SI), pase a la pregunta 3a

2a. ¿Ha sentido un deseo o necesidad tan fuerte de consumir _____ Droga de Impacto _____ que no pudo evitar hacerlo? 1 Sí 0 No

2b. ¿Ha deseado consumir _____ Droga de Impacto _____ tan desesperadamente que no podía pensar en nada más? 1 Sí 0 No

Dificultad para controlar el consumo

Después de la primera respuesta afirmativa (SI), pase a la pregunta 4a

3a. ¿Hubo ocasiones en que quiso suspender o disminuir el consumo de _____ Droga de Impacto _____ Si fue así ¿Ha sido siempre capaz de disminuir su uso por lo menos durante un mes? 1 Sí 0 No

3b. ¿Ha tenido períodos en los que usó _____ Droga de Impacto _____ en mayor cantidad o por más tiempo de lo que se proponía, o se le hizo difícil suspender el consumo, antes de sentirse intoxicado? 1 Sí 0 No

Abstinencia

Después de la primera respuesta afirmativa (SI), pase a la pregunta 5a

4a. En las horas o días siguientes a suspender o disminuir el uso de _____ Droga de Impacto _____ ¿Alguna vez tuvo malestares como temblores, sudores, no poder dormir, dolor de cabeza o estómago, etc.? 1 Sí 0 No

4b. ¿Utilizó _____ Droga de Impacto _____ u otra droga para evitar tener malestares como los que acaba de mencionar? 1 Sí 0 No

Reducción progresiva

Después de la primera respuesta afirmativa (SI), pase a la pregunta 6a

5a. ¿Ha habido ocasiones en que dedicaba mucho tiempo en conseguir _____ Droga de Impacto _____? 1 Sí 0 No

5b. ¿Ha pasado mucho tiempo consumiendo o recuperándose de los efectos de _____ Droga de Impacto _____? 1 Sí 0 No

5c. ¿Ha descontinuado o suspendido actividades importantes como estudios, deportes, trabajo, compartir con amigos o familiares, por conseguir o usar _____ Droga de Impacto _____? 1 Sí 0 No

Uso persistente

Se califica como síntoma presente si 6a y 6b = 1 ó 6c y 6d = 1

6a. ¿Ha tenido problemas de salud, como sobredosis accidental, los persistente, convulsiones, infecciones, hepatitis, abcesos, SIDA, problemas cardíacos u otra lesión relacionadas con el uso de _____ Droga de Impacto _____? 1 Sí 0 No

6b. ¿Continuó usando _____ Droga de Impacto _____ aun después de presentar estos problemas de salud? 1 Sí 0 No

6c. ¿Ha tenido problemas psicológicos o sociales asociados al uso de _____ Droga de Impacto _____ como sentirse deprimido, extraño o perseguido, o presentar fracasos laborales, conflictos familiares, actos de violencia, accidentes, etc.? 1 Sí 0 No

6d. ¿Continuó consumiendo _____ Droga de Impacto _____ aun después de saber que se relacionaba con alguno de estos problemas? 1 Sí 0 No

Figura 4.4: Cara trasera del formato de tamizaje de los años 2017-2019.

A) 2014 a 2016

B) 2017 a 2018

Figura 4.5: Formatos del tamizaje de los años A) 2014-2016 y B) 2017-2018.

```

1  Lectura de archivo.
2  -----
3  url = "dataset(7)_17-19.csv"
4  names = ['Edad', 'Sexo', 'Estado_Civil', 'Ocupacion', 'Estudio', 'Centro', 'Droga']
5  dataset = pandas.read_csv(url, names=names)
6
7
8  Configuración de métricas.
9  -----
10 array = dataset.values
11 X = array[:,0:6]
12 Y = array[:,6]
13 validation_size = 0.30
14 seed = 42
15
16 X_train, X_validation, Y_train, Y_validation =
17 model_selection.train_test_split(X, Y,
18 test_size=validation_size, random_state=seed)
19
20 scoring = 'accuracy'
21
22
23 Asignación de los algoritmos de IA a utilizar.
24 -----
25 models = []
26 models.append(('LR', LogisticRegression(solver='liblinear', multi_class='ovr')))
27 models.append(('KNN', KNeighborsClassifier()))
28 models.append(('CART', DecisionTreeClassifier()))
29 models.append(('NB', GaussianNB()))
30 models.append(('SVM', SVC(gamma='auto')))
31
32
33 Evaluación de los algoritmos de IA asignados.
34 -----
35 results = []
36 names = []
37 for name, model in models:
38     kfold = model_selection.KFold(n_splits=10, random_state=seed)
39     cv_results = model_selection.cross_val_score(model, X_train, Y_train,
40 cv=kfold, scoring=scoring)
41
42     results.append(cv_results)
43     names.append(name)
44     msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
45     print(msg)

```

Figura 4.10: Código de ejecución de las experimentaciones.

Capítulo 5

Análisis de resultados

En el presente capítulo se muestran los resultados obtenidos por las diferentes pruebas mencionadas con anterioridad, así como la de las validaciones, esto con el fin de conocer y comparar los incrementos y decrementos presentados en las diferentes pruebas.

5.1. Primeros Resultados

Los resultados obtenidos en las primeras experimentaciones mostraron un incremento al agregar nuevas características y al incrementar el volumen del *“dataset”*, sin embargo, a partir de la cuarta experimentación el porcentaje deja de mostrar mejoría, a pesar de que se continuó realizando pruebas agregando más características. Otro de los detalles encontrados al realizar las experimentaciones, fue que se determinó que la información referente a los años 2014 a 2016 presenta anomalías, pues al utilizar esta información y a pesar de tener un incremento de más del 50 % en los datos, no solo no se obtuvo un incremento, sino que todos los resultados presentaron un decremento, en comparación con las experimentaciones en las que sólo se utilizaron los años del 2017 a 2019. La tabla 5.1 presenta los resultados de las experimentaciones.

Cuadro 5.1: Resultados de las distintas experimentaciones con años 2014-2019

Experimentación	Años	Registros	% Precisión
Primera	17	6330	0.64
Segunda	17-18	13696	0.69
Segunda	14-18	36638	0.64
Tercera	17-18	13696	0.70
Tercera	14-18	36513	0.64
Cuarta	14-18	13696	0.71
Cuarta	14-18	36638	0.66
Quinta	17-18	13689	0.71
Quinta	14-18	36650	0.66
Sexta E1	17-18	13671	0.71
Sexta E1	14-18	36638	0.66
Sexta E2	17-18	13689	0.71
Sexta E2	14-18	36651	0.66
Sexta E3	17-18	13671	0.71
Sexta E3	14-18	36611	0.66
Séptima E1	17-18	13601	0.71
Séptima E1	14-18	36514	0.66
Séptima E2	17-18	13696	0.71
Séptima E2	14-18	36638	0.66
Octava	17-19	21003	0.71
Octava	14-19	41103	0.65
Novena	17-19	21003	0.72
Novena	14-19	41103	0.66
Iris	NA	150	0.99
Orange	17-18	13696	0.66
Orange	14-18	36638	0.61
WEKA	17-18	13696	0.71
WEKA	14-18	36638	0.66
WEKA2	17-18	13696	0.71
WEKA3	17-18	13696	0.71

La tabla 5.1 muestra los resultados obtenidos en todas las experimentaciones realizadas, tanto con el algoritmo desarrollado como las realizadas en software externos. La tabla incluye la experimentación, los años con los que se realizó el “*dataset*”, la cantidad de registros contenidos por el “*dataset*” y, finalmente, el porcentaje obtenido.

Los resultados reflejan el incremento previamente mencionado, de la misma manera, se presenta el resultado del software WEKA, el cual arrojó resultados similares a las experimentaciones con el algoritmo desarrollado; en cuanto al software de Orange presentó un decremento lineal en todas sus corridas, finalmente, se muestra el resultado de la validación, donde se utilizó la corrida clásica de la flor de Iris, obteniendo el porcentaje esperado, de 99 %.

5.2. Resultados utilizando validación cruzada

Como parte de las experimentaciones se realizaron cambios en el tipo de validación y el tipo de kernel utilizado en el algoritmo de SVM, la tabla 5.2 muestra los resultados obtenidos en este tipo de experimentación, en ella se presentan los años utilizados para la creación del “*dataset*”, también muestra cuatro columnas donde se describe cada uno de los porcentajes obtenidos, utilizando los diferentes tipos de kernel, el resultado se muestra tanto en el tiempo de procesamiento que requirió el algoritmo, como el porcentaje de precisión obtenido.

Cuadro 5.2: Tabla de resultados en Tiempo y Resultados validación cruzada.

Año	Linear	RBF	Sigmoid	Auto
2017	6s/51 %	8s/53 %	2s/51 %	9s/53 %
2018	40s/60 %	66s/64 %	13s/60 %	68s/64 %
2019	90s/60 %	140s/64 %	30s/61 %	132s/64 %
2014 a 2018	208s/46 %	373s/47 %	81s/46 %	377s/47 %
2017 a 2018 (centro)	16200s/NA	NA	NA	145s/76 %
2018 a 2019 (ciudad)	38s/72 %	60s/75 %	12s/72 %	60s/75 %
2017 a 2019 (ciudad)	85s/71 %	132s/74 %	29s/71 %	132s/74 %
2017 a 2019 (centro)	NA	NA	NA	333s/77 %
Iris	1.6s/97 %	0.4s/99 %	0.06s/34 %	0.8s/99 %

La tabla 5.2 muestra los resultados obtenidos con los diferentes tipos de kernel, en ella podemos observar que al utilizar el tipo de kernel de Sigmoid se obtiene un procesamiento mas rápido en comparación al resto, sin embargo, no presentó los mejores resultados en cuanto al porcentaje; en cambio, al utilizar el tipo RBF presentó mejores resultados, a pesar de que requirió más tiempo de procesamiento; en cuanto al tipo Auto, maneja resultados similares al tipo RBF, a pesar de tomar un poco más tiempo, no obtente, existen experimentaciones en las que el tipo RBF no logró concretar dicho procesamiento a pesar de utilizar días, mientras que al utilizar Auto se logró obtener un resultado.

En las experimentaciones realizadas con este tipo de validación se utilizaron principalmente datos de los años relacionados con el nuevo

formato de tamizaje, ya que estos presentaron un mejor resultados en las primeras experimentaciones, sin embargo, se optó por realizar por lo menos una experimentación con los 5 años, esto se realizó con la intención de descartar la posibilidad de obtener un mejor resultado utilizando específicamente un tipo de kernel; sin embargo, presentó restados aún más bajos y un elevado tiempo de procesamiento. Asimismo, se realizó una corrida con el “*dataset*” de la flor de Iris, para validar los resultados obtenidos con las diferentes fuentes de información, al obtener los resultados esperados, se concluyó que el utilizar este tipo de validación es viable.

El tipo auto fue el que presentó mejores resultados con respecto a los demás tipos de kernel, este logró incrementar su porcentaje de 53 a 77 %, como se puede observar en la tabla 5.2, este incrementó su resultado al utilizar los datos del año 2018, el cual obtuvo un 11 % más en comparación con el año 2017, sin embargo, mantuvo su resultado de 64 % al utilizar el año de 2019. Al realizar corridas con los años 2017 y 2018, este presentó un resultado de 76 %, finalmente, al utilizar los años de 2017,2018 y 2019, se obtuvo el resultado con mayor porcentaje de precisión, ya que este arrojó un 77 %.

Las pruebas realizadas donde se contempla el tiempo de procesamiento, se realizaron en un entorno controlado, es decir, se consideraron las condiciones del equipo para normalizar la carga de procesos externos al de entrenamiento.

Capítulo 6

Conclusiones y trabajo futuro

El presente capítulo presenta las conclusiones sobre el trabajo realizado, así como los resultados obtenidos. También se incluye la propuesta para trabajo futuro, ofreciendo un punto de partida para continuar robusteciendo y mejorando los resultados obtenidos.

6.1. Conclusiones

Al realizar un estudio profundo de la literatura y contar con el sustento de expertos en el tema, proporcionados por el Sistema de Vigilancia Epidemiológica de las Adicciones (SISVEA), se consiguió determinar las características que permitieron presentar los resultados obtenidos.

De igual manera, las pruebas realizadas con diferentes algoritmos de inteligencia artificial, permitió determinar el algoritmo de máquina de soporte vectorial como la mejor opción para el problema planteado en la presente investigación, obteniendo un 77 % de precisión.

Como parte de las pruebas realizadas, se demostró que al incrementar el número de registros aplicados al entrenamiento y validación, se obtuvo un incremento de los resultados, con lo cual se espera que al pasar los años e incrementar el número previamente mencionado, el algoritmo obtendrá un incremento en los resultados.

Los resultados obtenidos muestran que es posible diseñar e implementar un algoritmo para la detección y reconocimiento de patrones que presentan los pacientes que acuden a los centros de salud, los cuales pueden evolucionar a una sustancia con mayor riesgo.

Si bien el resultado obtenido al finalizar esta investigación, no es del todo satisfactorio, se encontró que la información recaudada no cumple con las 5Vs de big data, ya que ésta se ve comprometida, debido principalmente a que los pacientes son los que contestan la encuesta del tamizaje, por lo que las respuestas que ellos proporcionan, no reflejan necesariamente la realidad. Esto compromete la veracidad de esta información, la cual, es posible esté afectando los patrones reconocidos por el algoritmo, y que afecte directamente al resultado obtenido.

6.2. Trabajo futuro

Dentro de las perspectivas de trabajo futuro se incluye la implementación del algoritmo de detección y reconocimiento de patrones.

Este permitirá realizar un monitorio de los pacientes ingresados al sistema, y ayudará a los profesionales de la salud a tomar acciones pertinentes para la atención de éstos.

Como parte de los trabajos futuros es factible la realización de pruebas con un mayor número de datos, esto también incluye el considerar nuevos atributos que puedan surgir en las actualizaciones de

los formatos de tamizajes. Una actualización que mejoraría en gran medida la implementación, es el poder identificar al paciente a lo largo de su trayectoria por los centros de salud, esto con el fin de realizar perfiles de los pacientes y obtener una mayor precisión en los resultados, este tipo de implementación mejoraría en gran medida la veracidad de la información, sin embargo, esto no elimina del todo la inexactitud de la información.

Bibliografía

- [1] Martin Jelsma Guillermo Andrés ospina, Jorge Hernández tinajero. Amapola, opio y heroína la producción de colombia y méxico. 2018.
- [2] Chiang HL Chen Hsinchun, Storey Veda C. Analytics and algorithms, big data, cognitive computing, and deep learning in medicine and healthcare. *Mis Quarterly*, 36, 2012.
- [3] United Nations Office on Drugs and Crime. *United Nations Office on Drugs and Crime: World Drug Report 2019.*, volume 1. United Nations, 2019.
- [4] World Health Organisation. *Global Status Report on Alcohol and Health 2018.* World Health Organization, 2018.
- [5] Neurociencia del consumo y dependencia de sustancias psicoactivas, howpublished = <http://www.who.int/mediacentre/news/releases/2004/pr18/es/>, note = Accessed: 2018-11-24.
- [6] Velázquez Villatoro, Escobar Resendiz, Salazar Mujica, Cirett Bretón, Martínez Cañas, Hernández Soto, Ito Fregoso, Bautista Fleiz, Mora Medina, Reyes Gutiérrez, Núñez Franco, Martínez Romero, and Alvarado Mendoza. Encuesta nacional de consumo de drogas alcohol y tabaco (encodat) 2016_2017 (tabaco). *Reporte de Drogas , México*, 1, 2017.

- [7] Velázquez Villatoro, Escobar Resendiz, Salazar Mujica, Cirett Bretón, Martínez Cañas, Hernández Soto, Ito Fregoso, Bautista Fleiz, Mora Medina, Reyes Gutiérrez, Núñez Franco, Martínez Romero, and Alvarado Mendoza. Encuesta nacional de consumo de drogas alcohol y tabaco (encodat) 2016_2017 (tabaco). *Reporte de Tabaco*, México, 1, 2017.
- [8] Velázquez Villatoro, Escobar Resendiz, Salazar Mujica, Cirett Bretón, Martínez Cañas, Hernández Soto, Ito Fregoso, Bautista Fleiz, Mora Medina, Reyes Gutiérrez, Núñez Franco, Martínez Romero, and Alvarado Mendoza. Encuesta nacional de consumo de drogas alcohol y tabaco (encodat) 2016_2017 (alcohol). *Reporte de Alcohol*, México, 1, 2017.
- [9] Marcus Luciano de Oliveira Tavares, Amanda Márcia dos Santos Reinaldo, Eliana Aparecida Villa, Maria Odete Pereira, and Marlene Azevedo Magalhães Monteiro. Información, creencias y actitudes de escolares acerca del uso de alcohol y otras drogas. *SMAD. Revista eletrônica saúde mental álcool e drogas*, 15(2):45–51, 2019.
- [10] Ian Taylor. *Crime in context: A Critical Criminology of Market Societies*. Routledge, 2019.
- [11] Enrique López-Ramírez and Olga Inozemtseva. Validación del cuestionario de factores de riesgo interpersonales para el consumo de drogas en adolescentes (frida) en la población mexicana. *Cuadernos de Neuropsicología/Panamerican Journal of Neuropsychology*, 13(2), 2019.
- [12] G-Rubio-Valladolid J-Pascual-Arriazu. *Manual de drogodependencias para enfermería*. Ediciones Díaz de Santos, 2002.
- [13] Magaña-Alessandro-Rubio Monteverde-Horacio-Rubio. Breves comentarios sobre la historia del tabaco y el tabaquismo. *Re-*

- vista del Instituto Nacional de Enfermedades Respiratorias*, 19, 2006.
- [14] Ortiz-J-R Pérez-J-R Pérez-Cajaraville-J, Abejón-D. El dolor y su tratamiento a través de la historia. *Revista de la Sociedad Española del Dolor*, 12, 2005.
- [15] José Antonio Peña, Bustos Saldaña, and Mayra González Ruelas. Sustancias de abuso más frecuentes en México: aspectos médico-legales. *Gaceta internacional de ciencias forenses*, (32):9–28, 2019.
- [16] María-L Santiago-Castro. Adicción desde la hipótesis de la auto-medicación emocional: Una revisión. 2019.
- [17] Peter Szolovits. *Artificial intelligence in medicine*. Routledge, 2019.
- [18] Amit Konar. *Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain*. CRC press, 2018.
- [19] Shah-Dhavan-V McTavish-Fiona-Landucci-Gina Pe-Romashko Klaren-Gustafson David-H Kornfield-Rachel, Sarma-Prathusha-K. Detecting recovery problems just in time: Application of automated linguistic analysis and supervised machine learning to an online substance abuse forum. *Journal of medical Internet research*, 20, 2018.
- [20] Adriana Berenice Torres Valencia and Cristina Estrada Pineda. Structural vulnerability relational vulnerability drug dependent people stories. *Revista Electrónica de Psicología Iztacala*, 22(1):293–318, 2019.
- [21] Becoña-Iglesias-Elisardo. Bases psicológicas de la prevención del consumo de drogas. *Papeles del Psicólogo*, 28, 2007.

- [22] Consumo de drogas inyectables y vih. <https://www.who.int/hiv/topics/idu/about/es/>. Accessed: 2018-11-23.
- [23] Pancardo-Rafael-Villalobos-Vázquez-Myrna-Miriam-Valera-Mota-José Luis Cadena Anguiano-Juan Francisco Estrada-Olivia del Rocío Guzmán Astrid-Gabriela Maldonado Bonilla Dulce María Luisa Ortiz Molina-Diana, Cecilia-Tapia. *Adicciones En El Adolescente*. Universidad Nacional Autónoma de México, 2016.
- [24] Gloria S. Rubiano Ordoñez Jorge A. Avellaneda Rincón Germán D. Sicard Arenas. Dependencia a Consumo de Sustancias Psicoactivas. Una mirada comparativa desde la Psicología Clínica Sistémica y Clínica Comportamental. PONTIFICIA UNIVERSIDAD JAVERIANA. Facultad de Psicología, Noviembre 2015.
- [25] Organización Mundial de la Salud. Glosario de términos de alcohol y drogas . https://www.who.int/substance_abuse/terminology/lexicon_alcohol_drugs_spanish
- [26] Definición de adicción a sustancias o drogas. <http://www.imss.gob.mx/salud-en-linea/adicciones>. Accessed: 2018-10-16.
- [27] Landivar-Universidad Rafael. Adicciones ¿qué es una adicción? *Universidad Saludable*, 1, 2014.
- [28] Redondo Romero M.I. Hidalgo Vicario M.I. Consumo de drogas en la adolescencia. 2, 2005.
- [29] P San Juan Sanz. Protocolo diagnóstico y tratamiento del síndrome de abstinencia. *Medicine-Programa de Formación Médica Continuada Acreditado*, 12(85):5004–5008, 2019.
- [30] Peter Janik, Michaela Kosticova, Jan Pecenek Prof, and Michal Turcek. Categorization of psychoactive substances into “hard

- drugs” and “soft drugs”: a critical review of terminology used in current scientific literature. *The American Journal of Drug and Alcohol Abuse*, 43(6):636–646, 2017.
- [31] Maria del Pilar Navia-Bueno, Jacqueline Farah-Bravo, Nina Yaksic-Feraudy, Patrícia Philco-Lima, and Angela Maria Magosso Takayanagui. Knowledge on drugs phenomenon by students and faculty from the medical school at universidad mayor de san andrés, la paz, bolivia. *Revista latino-americana de enfermagem*, 19(SPE):722–729, 2011.
- [32] CAT Barcelona. ¿qué es el alcohol y qué efectos produce? <https://www.cat-barcelona.com/faqs/view/que-es-el-alcohol-y-que-efectos-produce>, 2008. Online; Accesado 05 Febrero 2019.
- [33] Dilsen Colak, Arzu Oguz, Dogan Yazilitas, Inanc Goksen Imamoglu, and Mustafa Altinbas. Morphine: patient knowledge and attitudes in the central anatolia part of turkey. *Asian Pac J Cancer Prev*, 15(12):4983–4988, 2014.
- [34] I Gainza, S Nogué, C Martínez Velasco, RS Hoffman, G Burillo-Putze, A Dueñas, J Gómez, and MA Pinillos. Intoxicación por drogas. In *Anales del Sistema Sanitario de Navarra*, volume 26, pages 99–128. SciELO Espana, 2003.
- [35] Alejandra E Ruiz Contreras, Mónica Méndez Díaz, Bertha Prieto Gómez, Antonio Romano, Seraid Caynas, and Oscar Prospéro García. El cerebro, las drogas y los genes. *Salud mental*, 33(6):535–542, 2010.
- [36] Katherine Calvo Mata. *Facultad de Medicina Escuela de Enfermería*. PhD thesis, Universidad de Costa Rica, 2015.
- [37] Organización Mundial de la Salud. Tabaco. <https://www.who.int/es/news-room/fact-sheets/detail/tobacco>. Online; Accesado 08 Febrero 2019.

- [38] Allister Vale. Drugs of abuse (amfetamines, bzp, cannabis, cocaine, ghb, lsd). *Medicine*, 40(2):84–87, 2012.
- [39] Elisardo Becoña Iglesias and Maite Cortés Tomás. Manual de adicciones para psicólogos especialistas en psicología clínica en formación, 2016.
- [40] David E Nichols and Charles S Grob. Is lsd toxic? *Forensic science international*, 284:141–145, 2018.
- [41] Josep Solé Puig. Lsd y alucinógenos. *Adicciones*, 15(5):179–198, 2003.
- [42] MI Colado. Éxtasis (mdma) y drogas de diseño: estructura, farmacología, mecanismos de acción y efectos en el ser humano. *Trastornos adictivos*, 10(3):175–182, 2008.
- [43] Ramos Lira Luciana Caballero Gutiérrez Miguel Angel. Violencia: Una revisión del tema dentro del marco de trabajo de investigación en el instituto nacional de psiquiatría. *Salud Mental*, 27, 2004.
- [44] Jenniffer Lisseth Baños Cedeño. Violencia intrafamiliar y la salud mental en los estudiantes de la unidad educativa siete de octubre. B.S. thesis, BABAHOYO: UTB, 2018, 2018.
- [45] Meredith Linette Cedeño Granda. Violencia intrafamiliar y su incidencia en el rendimiento escolar de los estudiantes del cuarto año de educación general básica de la unidad educativa ecuador, de la parroquia san carlos, cantón quevedo. B.S. thesis, BABAHOYO: UTB, 2018, 2018.
- [46] José Luis Rivera Salazar, Verónica Ramírez Salgado, Ana Laura Rojas Patiño, Claudia Cecilia Vega Robles, and Mayra Rodríguez. *Violencia autoinfligida y la adicción al dolor*, pages 67–86. 01 2016.

- [47] Mauricio Olavarria-Gambi. Delito y violencia. causas, consecuencias y estrategias de prevención. *Documentos de Apoyo Docente (DAD)*, N^o. 20, 2007/106 pags., 01 2007.
- [48] Etienne Krug, James Mercy, L.L. Dalhberg, and Anthony Zwi. El informe mundial sobre la violencia y la salud. *Biomédica*, 22:326–336, 12 2002.
- [49] Halper Fern Kaufman Marcia Hurwitz Judith S, Nugent Alan. *Big data for dummies*. John Wiley Sons, 2013.
- [50] Soche López Sebastián. Metodología para el modelamiento de datos basado en big data, enfocados al consumo de tráfico (voz-datos) generado por los clientes. 2016.
- [51] de Abajo Martínez Nicolás Díez Raúl Pino, Gómez Alberto Gómez. *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva*. Universidad de Oviedo, 2001.
- [52] García Alberto. *Inteligencia artificial: fundamentos, práctica y aplicaciones*. Rc Libros, 2012.
- [53] Ravindra Kumar and Saloni Garg. Aspect-based sentiment analysis using deep learning convolutional neural network. In *Information and Communication Technology for Sustainable Development*, pages 43–52. Springer, 2020.
- [54] Antonio Diez Patricio. More on interpretation (i). reasoning and truth. (6):363–382, 2016.
- [55] Revista digital para profesionales de la enseñanza. Temas para la educación. (2):6, 2009.
- [56] Olga Marino. Representacion del conocimiento: Aportes a la informatica educativa. 02 2019.

- [57] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016.
- [58] John J Sanabria and John F Archila. Detección y análisis de movimiento usando visión artificial. *Scientia et technica*, 3(49):180–188, 2011.
- [59] Haider Murtaza Gandomi Amir. Beyond the hype: Big data concepts, methods, and analytics. 35, 2015.
- [60] Consuelo Alejandra Bello Zamorano. *Análisis predictivo de pacientes No-Show mediante un modelo multivariable basado en fuentes de información heterogéneas*. PhD thesis, Universidad Andrés Bello, 2018.
- [61] Carlos Espino Timón. Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas open source que permiten su uso. 2017.
- [62] José Miguel Vegas Valle et al. Diseño e implementación de nuevas herramientas para el análisis de la prevalencia e incidencia de diabetes mellitus. 2019.
- [63] Soteris A Kalogirou. Artificial neural networks in renewable energy systems applications: a review. *Renewable and sustainable energy reviews*, 5(4):373–401, 2001.
- [64] Culquichicón Vílchez and Luis German. Análisis comparativo de métodos de reconocimiento facial en plataforma android para resolver problemas de seguridad e iluminación en ambientes no controlados. 2018.
- [65] Enrique J Carmona Suárez. Tutorial sobre máquinas de vectores soporte (svm). *Tutorial sobre Máquinas de Vectores Soporte (SVM)*, 2014.

- [66] Rubén Solera Ureña. Máquinas de vectores soporte para reconocimiento robusto de habla. 2011.
- [67] Tomás Santibáñez-Arellano. *Validez predictiva de un clasificador basado en máquina de soporte vectorial para éxito o fracaso en la extubación de pacientes conectados a ventilación mecánica invasiva en una unidad de paciente crítico adulto*. PhD thesis, 09 2015.
- [68] Pablo Vinues. Regresión lineal simple y polinomial: teoría y práctica. 1, 2016.
- [69] Beatriz Molina Serrano, Nicoletta González-Cancelas, and Francisco Soler-Flores. Gestión de la sostenibilidad portuaria basada en un modelo de redes bayesianas. aplicación al sistema portuario español. *Ingeniare. Revista chilena de ingeniería*, 26(4):631–644, 2018.
- [70] Jia Ping You Jane Liu Xiaofeng, Vijaya Kumar B. V.K. Hard negative generation for identity-disentangled facial expression recognition. 2018.
- [71] Teodoro Calonge Cano Luis Alonso Romero. Capitulo 1.-redes neuronales y reconocimiento de patrones. 2001.
- [72] JP Connor, M Symons, GFX Feeney, R McD Young, and Janet Wiles. The application of machine learning techniques as an adjunct to clinical decision making in alcohol dependence treatment. *Substance use & misuse*, 42(14):2193–2206, 2007.
- [73] Elisa Capecci, Nikola Kasabov, and Grace Y Wang. Analysis of connectivity in neucube spiking neural network models trained on eeg data for the understanding of functional changes in the brain: A case study on opiate dependence treatment. *Neural Networks*, 68:62–77, 2015.

- [74] John Haskell. Why we created triggr health. <https://blog.triggrhealth.com/why-we-created-triggr-health-64e9701cffd1>, 2017. Accessed: 2019-02-28.
- [75] Tao Ding, Warren K Bickel, and Shimei Pan. Social media-based substance use prediction. *arXiv preprint arXiv:1705.05633*, 2017.
- [76] P Kruchten. Architectural blueprints—the 4+ 1 view model of software architecture, software, vol. 12, number 6, 1995.
- [77] Cristina Venera Geambaşu. Bpmn vs uml activity diagram for business process modeling. *Accounting and Management Information Systems*, 11(4):637–651, 2012.
- [78] Grady Booch, James Rumbaugh, Ivar Jacobson, José Sáez Martínez, and Jesús J García Molina. *El lenguaje unificado de modelado*, volume 1. Addison wesley Madrid, 1999.
- [79] Secretaría de salud. Informe SISVEA 2016. Technical report, Dirección general de epidemiología, Junio 2016.
- [80] Yiming Zhang, Yujie Fan, Yanfang Ye, Xin Li, and Erin L Winstanley. Utilizing social media to combat opioid addiction epidemic: Automatic detection of opioid users from twitter. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [81] National Institute on Drug Abuse. Overdose death rates. <https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates>. Accessed: 2019-03-04.