



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

INSTITUTO DE CIENCIAS BÁSICAS E INGENIERÍA

HERRAMIENTA PARA EL ANÁLISIS DE FACTORES
DE RIESGO EN CASOS DE SUICIDIO EN
POBLACIONES GERIÁTRICAS USANDO
INTELIGENCIA COMPUTACIONAL

T E S I S

Maestría en Ciencias en Computación Avanzada y
Electrónica

PRESENTA:

Alondra Sánchez Salinas

DIRECTORA:

Dra. Anilu Franco Arcega

COMITÉ TUTORIAL:

Dr. Virgilio López Morales

Dr. Esteban Rueda Soriano

Dr. Manuel Alejandro Ojeda Misses

Mineral de la Reforma, Hidalgo, 2024

A mis abuelos, cuyo cariño y legado se ven reflejados en la persona en la que me he convertido.

Agradecimientos

Mi mayor gratitud es para Dios, por cuidarme y bendecirme poniendo en mi camino a personas maravillosas que me ayudaron a alcanzar el logro más importante en mi vida académica hasta el momento.

A mi familia, quien me ha dado su amor, apoyo, sabiduría y protección incondicionalmente, sin los cuales no habría llegado hasta la culminación de este trabajo.

A mis amigos, instructores e ídolos que me ayudaron a crecer y convertirme en una mejor persona y profesionista de lo que era cuando empecé este recorrido.

A los docentes que con su amor por la excelencia me apoyaron para que continuara con el programa. Especialmente a mi directora de tesis, la Dra. Anilu Franco Arcega, que tuvo fe en mí y con paciencia, me apoyó en momentos de adversidad, sin mencionar que fue un pilar importante en el desarrollo de este trabajo.

A la Dra. Lilian Elizabeth Bosques Brugada y a la doctoranda Christian Jazmín Flores Pérez, quienes fueron el corazón de este proyecto y me dieron la oportunidad de colaborar en su investigación con la esperanza de ayudar a que nuestra sociedad avance a un estado en el que la vida de nuestros adultos mayores sea debidamente protegida y apreciada.

Finalmente, agradezco al Consejo de Ciencia y Tecnología (CONACYT) por su comprensión, apoyo y el financiamiento de la beca a la cual fui acreedora para llevar a cabo todas las actividades que dieron como fruto esta tesis.

Resumen

Según el Instituto Nacional de Estadística y Geografía (INEGI), la cantidad de suicidios en México ha ido continuamente en aumento. Siendo así, que tan solo en el año 2012 se registró un total de 5,549 muertes a causa de lesiones autoinfligidas y para el año 2021, fueron 8,432 decesos totales por esta causa. Esto implica aproximadamente la pérdida de 23 personas al día durante ese último año.

La aplicación de las ciencias computacionales, específicamente del área de la Inteligencia Artificial ha logrado resolver un sinnúmero de problemas de la vida diaria de hoy en día. Es acertado pensar que esta herramienta computacional y sus diversas ramas pueden ser usadas para ayudar a entender mejor el problema del suicidio y probablemente, apoyar a su resolución.

En este proyecto, se pretende desarrollar una herramienta computacional que permita la adquisición, organización y procesamiento de información relevante acerca del suicidio en la población geriátrica y que sirva de apoyo a los especialistas en el área para detectar si un paciente es vulnerable a cometer autolesiones y de ese modo, hacer una intervención o proporcionar un tratamiento adecuado para evitar el deceso por propia mano de la persona en cuestión.

Índice general

Índice de figuras	v
Índice de tablas	vi
Introducción	1
1. Marco Teórico	9
1.1. Suicidio	9
1.1.1. Ideación suicida	10
1.1.2. Factores de riesgo	10
1.1.2.1. Factores sociales	11
1.1.2.2. Factores personales	12
1.1.2.3. Factores psicológicos	13
1.1.2.4. Factores precipitantes	14
1.1.2.5. Trastornos mentales, como factor de riesgo	15
1.1.3. Factores protectores	15
1.1.3.1. Individuales	16
1.1.3.2. Familiares	16
1.1.3.3. Comunitarios	17
1.1.3.4. Institucionales	18
1.1.4. Panorama actual del suicidio en México	18
1.2. Inteligencia Artificial	20
1.2.1. Inteligencia Computacional	21
1.2.2. Reconocimiento de patrones	22
1.2.3. Minería de Datos	23
1.2.4. Aprendizaje Automático	25
1.3. Tipos de aprendizaje	26
1.3.1. Aprendizaje supervisado	27
1.3.2. Aprendizaje no supervisado	29
1.3.3. Aprendizaje semisupervisado	31
1.4. Metodologías	32
1.4.1. Tradicionales	32
1.4.1.1. Cascada	33

1.4.1.2.	Espiral	34
1.4.1.3.	Incremental	35
1.4.2.	Ágiles	37
1.4.2.1.	Kaizen	37
1.4.2.2.	Kanban	39
2.	Herramientas Utilizadas	42
2.1.	Herramientas de desarrollo	42
2.1.1.	Python	42
2.1.2.	Visual Studio Code	43
2.1.3.	StarUML	43
2.1.4.	MockFlow	43
2.1.5.	Flask	44
2.1.6.	JavaScript	44
2.1.7.	PosgreSQL	44
2.1.8.	Microsoft Azure	45
2.2.	Algoritmos usados	46
2.2.1.	K-Means	46
2.2.2.	Método del codo	46
2.2.3.	Inercia	47
2.2.4.	Análisis de componentes principales (PCA)	47
2.2.5.	Regresión logística	48
2.2.5.1.	Log-Likelihood	48
2.2.5.2.	Pseudo R^2	49
2.2.5.3.	Grados de libertad del modelo (Df_model)	49
2.2.5.4.	LL-Null	49
2.2.5.5.	LLR p-value	50
2.2.5.6.	Coefficientes estimados (Coef)	50
2.2.5.7.	Error estándar (std err)	51
2.2.5.8.	Valor de Z	51
2.2.5.9.	$P > z $	51
2.2.5.10.	Límites del intervalo de confianza	52
2.2.5.11.	Likelihood Ratio Chi-square	52
2.2.6.	Regresión lineal	53
2.2.6.1.	Coefficiente de determinación R^2	53
2.2.6.2.	R^2 ajustado (Adj. R-squared)	54
2.2.6.3.	Estadístico F (F-statistic)	54
2.2.6.4.	Prob (F-statistic)	55
2.2.6.5.	Akaike Information Criterion (AIC)	55
2.2.6.6.	Bayesian Information Criterion (BIC)	56
2.2.6.7.	Estadístico t	56
2.2.6.8.	$P > t $	57
2.3.	Herramientas de análisis estadístico de datos	57

ÍNDICE GENERAL

2.3.1. Máximo y mínimo	57
2.3.2. Media aritmética	58
2.3.3. Desviación estándar	58
2.3.4. Rango intercuartílico	59
2.3.5. Prueba de bondad de ajuste de Kolmogórov-Smirnov	59
2.3.6. Correlación lineal	60
3. Estado del Arte	61
3.1. Un sistema informatizado de apoyo a las decisiones clínicas para el manejo de la depresión en atención primaria	61
3.2. How Facebook AI Helps Suicide Prevention	62
3.3. Suicide and Suicide Attempt Descriptors by Multimethod Approach	63
3.4. Recognizing states of psychological vulnerability to suicidal behavior: a Bayesian network of artificial intelligence applied to a clinical sample	65
3.5. Suicide risk configuration system in a clustered clinical sample: a generalized linear model obtained through the LASSO technique	66
3.6. Identification of Suicide Attempt Risk Factors in a National US Survey Using Machine Learning	67
3.7. A predictive model for automatic detection of loneliness and social isolation using machine learning	68
3.8. Physical frailty prediction model for the oldest old	68
3.9. Discusión	69
4. Metodología de la Investigación	71
4.1. Planificar	72
4.2. Ejecutar	72
4.3. Probar	73
5. Planeación de la Construcción de la Herramienta	75
5.1. Análisis de requerimientos	75
5.2. Diseño de la base de datos	76
5.3. Descripción de los procesos en el sistema	78
5.3.1. Diagrama Casos de uso	80
5.4. Wireframe	81
5.4.1. Inicio de sesión	81
5.4.2. Menú	82
5.4.3. Alta de adulto mayor	83
5.4.4. Cuestionarios	83
5.4.5. Resultados	84
5.4.6. Registro de nuevos usuarios	84
5.4.7. Tabla de estadísticas básicas	85
5.4.8. Sección de análisis de datos	85

6. Desarrollo de la Herramienta Propuesta	87
6.1. Implementación de un servidor web local para las pruebas	87
6.2. Construcción de la base de datos	88
6.3. Desarrollo de la aplicación web	89
6.3.1. Inicio de sesión	90
6.3.2. Menú principal	91
6.3.3. Alta de datos y consentimiento informado	92
6.3.4. Recolección de datos faltantes	93
6.3.5. Datos Sociodemográficos	93
6.3.6. InBody	94
6.3.7. Cuestionarios	95
6.3.8. Editar y eliminar	95
6.3.9. Resultados	97
6.3.10. Registro de nuevos usuarios	98
6.3.11. Tabla de estadísticas básicas	99
6.3.12. Sección de análisis de datos	100
6.4. Instalación del sistema en un servidor en la nube	101
6.4.1. Creación del servidor de base de datos	102
6.4.2. Creación del servidor web en la nube	103
6.4.3. Instalación del sistema en el servidor web	103
7. Validación de la Herramienta	105
7.1. Evaluación de requerimientos	105
7.2. Evaluación de Sistema de Escalas de Usabilidad	106
Conclusiones	109
7.2.1. Aportaciones	109
7.2.2. Trabajo futuro	110
Referencias	111
A. Consentimiento informado para persona adulta mayor y responsable	120
B. Instrumentos de medición	122
B.1. Cuestionario Autoestima	122
B.2. Cuestionario Gratitud	122
B.3. Cuestionario Apreciación corporal	123
B.4. Cuestionario Ansiedad y Depresión	124
B.5. Cuestionario Ideación suicida	125
B.6. Cuestionario Aprecio Funcional	125
B.7. Cuestionario Lubben	126

Índice de figuras

1.1. Tasa de suicidios según sexo (Por cada 100 mil habitantes) Fuente: INEGI (2022)	19
1.2. Tasa de suicidio en la población de 15 a 29 años desde 2015 hasta 2021 (Por cada 100 mil habitantes))(1)	19
1.3. Porcentaje de la población de 10 años y más que ha pensado suicidarse, por sexo y grupo de edad 2018))(1)	20
1.4. Modelo de Minería de Datos (Elaboración propia basada en (2))	24
1.5. Ejemplo de un conjunto de entrenamiento etiquetado para aprendizaje supervisado (filtro de spam)(3)	28
1.6. Metodología de Cascada (Elaboración propia)	33
1.7. Metodología de Espiral (Elaboración propia basada en (4))	35
1.8. Metodología incremental (Elaboración propia)	36
1.9. Metodología Kaizen (Elaboración propia)	38
1.10. Metodología de Kanban (Elaboración propia)	40
4.1. Metodología Kaizen-Kanban propuesta (Elaboración propia)	71
5.1. Diagrama entidad-relación	77
5.2. Diagrama de procesos almacenados y disparadores en una BD	78
5.3. Diagrama de Casos de Uso del sistema a desarrollar (Elaboración propia)	81
5.4. Wireframe del inicio de sesión del sistema (Elaboración propia)	82
5.5. Wireframe del menú del sistema (Elaboración propia)	82
5.6. Wireframe del formulario para ingresar los datos del adulto mayor (Elaboración propia)	83
5.7. Wireframe de los cuestionarios (Elaboración propia)	84
5.8. Wireframe de los resultados (Elaboración propia)	84
5.9. Wireframe de los registro de usuarios (Elaboración propia)	85
5.10. Wireframe de la tabla de estadísticas básicas	86
5.11. Wireframe de la sección de análisis de datos	86
6.1. Carpeta con Python y Pip instalados	88
6.2. Ejecución del servidor	88
6.3. Interfaz para el inicio de sesión del sistema	90

6.4. Menú principal para administradores	91
6.5. Menú principal para pasantes	92
6.6. Formulario para el alta de datos de un paciente	93
6.7. Apartado para firmar digitalmente el consentimiento informado	94
6.8. Generación del archivo PDF	95
6.9. Tabla en el sistema	96
6.10. Interfaz para recolectar los datos sociodemográficos de los pacientes	96
6.11. Interfaz para recolectar los datos de la prueba InBody de los pacientes	97
6.12. Cuestionario a contestar por el adulto mayor	97
6.13. Menú desplegado de opciones para editar y eliminar un registro	98
6.14. Tabla en la que se muestran los resultados del paciente y los valores de referencia	98
6.15. Sección en la que se registran los nuevos usuarios	99
6.16. Sección en la que se presenta la tabla de estadísticas básicas	99
6.17. Sección en la que se realiza la prueba de Kolmogorov-Smirnov	100
6.18. Sección en la que se realiza el agrupamiento de los datos con K-Means	101
6.19. Sección en la que se aprecia la gráfica de grupos generados por K-Means y la tabla de caracterización de los centroides	101
6.20. Vista de Azure para la configuración del servicio del servidor de base de datos	102
6.21. Vista de Azure para la configuración del servicio del servidor web	103
6.22. Vista de Azure para la configuración del servicio del servidor de base de datos	104
6.23. Vista de Azure para la configuración del servicio del servidor de base de datos	104

Índice de tablas

3.1. Tabla comparativa de trabajos relacionados	70
7.1. Tabla de la evaluación de requerimientos por parte de los especialistas .	106
7.2. Tabla de las evaluaciones SUS de los especialistas que usaron el sistema	108

Introducción

Hoy en día, los humanos han perdido casi por completo el instinto natural que muchos animales aún conservan. Que en general, contribuye a la preservación de la vida como individuo y especie, esto es debido a que las grandes sociedades, cadenas de comercio y servicios establecidos que se han construido, le han proveído las comodidades necesarias para dejar de lado los problemas que son de principal preocupación para el resto de las especies.

Sin embargo, esto no representa el final de los problemas de la humanidad, por el contrario, lo que pareciera ser una ventaja evolutiva a su favor, ha terminado por convertirse en una de las herropeas más pesadas para su supervivencia. Desde la antigüedad, la especie humana ha atravesado problemas mucho más complejos que el resto de los seres vivos, en particular los de índole social y religioso. Siendo así que continuamente se visualizaban escenarios donde las adversidades sobrepasaban sus capacidades para superarlas y la muerte se presentaba como última alternativa.

Es por ello que el suicidio fue aceptado en las sociedades primitivas como una solución práctica e incluso honorable al enfrentar un problema. Por ejemplo, en la prehistórica Mesoamérica, donde ciertos grupos optaban por inmolarsse para complacer a sus deidades o en el Japón medieval, donde los samuráis practicaban el seppuku, que se trataba de un ritual donde el guerrero terminaba con su vida, con el propósito de no aceptar la derrota.

La Organización Mundial de la Salud, define el suicidio como un acto deliberada-

mente iniciado y realizado por una persona en pleno conocimiento o expectativa de su desenlace fatal. Así mismo, advierte que se trata de un problema multifactorial, que resulta de una compleja interacción de factores biológicos, genéticos, psicológicos, sociológicos y ambientales. No obstante, también infiere que la prevención y el tratamiento adecuado de las crisis y trastornos provocados por los factores anteriormente mencionados, permite la reducción de las tasas de suicidio.

Aunado a todo lo anterior, las nuevas tecnologías han logrado resolver un sinnúmero de dificultades contemporáneas y cotidianas, apoyándose de diversas disciplinas, como la Inteligencia Artificial, la cual abarca entre sus ramas más destacadas, a la Minería de Datos, Inteligencia Computacional, Aprendizaje Profundo, Aprendizaje Automático, Reconocimiento de Patrones, entre otros.

Los sistemas informáticos suelen ser muy útiles al momento de sustentar un proceso de toma de decisiones, lo que implica el uso de datos y modelos para la generación, estimación, evaluación y/o la comparación sistemática de alternativas. Estos resultados han permitido ayudar a los especialistas a reunir inteligencia, generar opciones y tomar decisiones para la resolución de un problema determinado. Así pues, estos sistemas también se emplean para obtener información que revele los elementos destacados de los problemas y las relaciones entre ellos, así como para identificar, crear y comunicar las formas de proceder disponibles.

Entre los ejemplos más destacados de cómo la Inteligencia Artificial ha mejorado la calidad de vida humana en sectores como el comercio, la logística y la salud, se encuentra e-commerce, que a través del procesamiento de los datos del comprador, ofrece una atención personalizada, aumentando la probabilidad de compra. Uno más, podría ser la asistencia de un Sistema de Posicionamiento Global, también conocido como GPS por sus siglas en inglés, que ofrece la posibilidad de saber cual es la localización exacta de una persona o la ruta más rápida para llegar a un determinado lugar.

Por último, en el ámbito de la salud, se han diseñado varios sistemas, los cua-

les dependiendo de qué tan precisos sean los datos proporcionados, podrían hacer un diagnóstico de alguna enfermedad que sufra un paciente y ofrecer un tratamiento oportuno.

En particular, en este trabajo se aborda una problemática social latente en nuestros días, proponiendo el uso de diferentes técnicas de Inteligencia Computacional para descubrir conocimiento en la información disponible de casos de suicidio, dado el constante aumento de casos de esta índole en México.

Planteamiento del problema

Factores de tipo biológicos, genéticos, psicológicos, sociológicos y ambientales, tales como el género, la raza, trastornos psicológicos, estrés laboral o la constante exposición a entornos que producen agobio, contribuyen a que la población sea vulnerable a buscar un desenlace fatal de su vida para terminar con sus problemas.

Según el Instituto Nacional de Estadística y Geografía (INEGI), la cantidad de suicidios en México ha ido continuamente en aumento. Siendo así, que tan solo en el año 2012 se registró un total de 5,549 muertes a causa de lesiones autoinfligidas y para el año 2021, fueron 8,432 decesos totales por esta causa. Esto implica aproximadamente la pérdida de 23 personas al día durante ese último año.

A pesar de los esfuerzos de los diferentes organismos que luchan por prevenir que las personas vulnerables a cometer suicidio tomen la decisión de terminar con su vida, este problema sigue existiendo y empeorando.

Además, al tratarse de un tema que aún sigue siendo un tabú para la sociedad en general, se ha preferido ignorar el tema y dar pie a que se generen muchos mitos al respecto y desinformación, agravando la situación. Sin mencionar que se trata de una práctica repudiada por las sociedades religiosas de todo el mundo, por lo que pueden desconocer a uno de sus miembros si este optó por quitarse la vida. Uno de los sectores

más vulnerables a este problema es la población geriátrica, debido a que la calidad de vida disminuye en diferentes aspectos. A consecuencia de este hecho y muchos otros, la cifra documentada de los adultos mayores que se han quitado la vida, no es muy precisa.

Es por eso que se ha tomado la iniciativa de recopilar y analizar datos al respecto en conjunto con especialistas en el tema, con el fin de llevar a cabo un estudio más detallado de los factores que podrían estar relacionados con esa causa de muerte en adultos mayores. Además de estudiar la información disponible que brinda el INEGI sobre personas que han cometido suicidio.

Con el objetivo de apoyar a especialistas en el tratamiento de esta tipo de información, se propone desarrollar una herramienta que permita gestionar datos relacionados a diversos factores protectores y de riesgo que pudieran llevar a una persona a atentar contra su vida. Así mismo, se propone hacer uso de herramientas de Inteligencia Computacional que faciliten el proceso y encuentren información relevante, de modo que un problema que es real y ha proliferado en estos tiempos, sea atendido por herramientas actuales y oportunas, aún cuando es un problema que no tiene su origen en la actualidad.

Justificación

En México, son pocos los organismos que se encargan de la prevención del suicidio, además de que no se dedican especialmente a ese problema, sino que están sobrecargados por el déficit de especialistas que atienden trastornos relacionados con dicha causa de muerte, como depresión, esquizofrenia y abuso de sustancias adictivas.

Los adultos mayores son una población más vulnerable debido a la pérdida paulatina de las funciones físicas y/o mentales, por lo que difícilmente logran obtener la ayuda que necesitan.

Los especialistas tratan de atender la mayor cantidad de personas con ideología suicida, sin embargo, sus procedimientos son obsoletos ya que, la recopilación de información en diversas ocasiones se hace a través de lápiz y papel, lo que resulta en procesos tediosos y vulnerables a errores. Además, encontrar relaciones relevantes entre los datos que describen la percepción de la población geriátrica sobre la ideación suicida y los casos consumados de lesiones autoinfligidas, implementando técnicas de Inteligencia Computacional, sería importante para ayudar a estos organismos a diseñar mejores estrategias que puedan ser dirigidas al sector correcto de la población y este tenga acceso a información oportuna que los aliente a desistir de la opción de suicidarse.

De este modo, se estaría ahorrando capital humano al delegarle parte del trabajo al área computacional y quienes necesiten de ese conocimiento para una efectiva toma de decisiones, puedan obtenerlo de una forma más sencilla y rápida.

Aunado a lo anterior, como se verá en secciones posteriores, la mayoría de los investigadores de las áreas involucradas no se han concentrado en desarrollar una herramienta con los propósitos y recursos descritos en esta investigación.

Hipótesis

Considerando el planteamiento anterior, en este trabajo se generan dos preguntas de investigación:

- ¿Se puede mejorar el proceso de adquisición de datos para contar con información de adultos mayores en una forma estructurada y homogénea?
- ¿Es posible descubrir conocimiento a partir de esta información que apoye la prevención del suicidio en poblaciones geriátricas?

Partiendo de estas preguntas se define la siguiente hipótesis.

Contar con una herramienta computacional que permita la adquisición, organización y procesamiento de información relevante en esta problemática social, apoyará el

descubrimiento de conocimiento en conjuntos de datos que integren casos de adultos mayores.

La herramienta propuesta permitirá conocer el impacto que tienen los factores personales, sociales, familiares, entre otros, en la decisión de optar por dicha acción. Así como también, determinar las características más comunes entre quiénes pertenecen a la población más vulnerable a cometer autolesiones que terminen con su vida.

Objetivos

Con la finalidad de cumplir con el propósito de este trabajo, se establece el objetivo general a cumplir, así como los objetivos específicos que permitirán completarlo.

Objetivo General

Desarrollar una herramienta computacional para el descubrimiento de conocimiento en información de casos de suicidio en adultos mayores, implementando técnicas de análisis estadístico y de Inteligencia Computacional.

Objetivos específicos

- Integrar la información relacionada con el caso práctico abordado en el trabajo, analizando repositorios de datos poblacionales, para establecer el marco de referencia del trabajo.
- Implementar diversos algoritmos y estrategias de análisis que permitan el procesamiento de datos para encontrar relaciones valiosas entre ellos.
- Diseñar el software que gestione los datos relacionados al caso de estudio, aplicando las fases de una metodología de desarrollo para la construcción de la herramienta.

-
- Validar la herramienta desarrollada a través de un instrumento de evaluación para determinar su utilidad.

Estructura del documento

La presente tesis pretende mostrar el trabajo realizado para la construcción del sistema con el que se planteó resolver la problemática descrita en esta introducción. Para tal fin, se estructura el documento de la siguiente manera:

- Capítulo 1, Marco Teórico describe los principales conceptos que serán empleados en el desarrollo de esta investigación. Dichos conceptos refieren a tres vertientes principales: el problema social abordado, el conjunto de disciplinas que otorgan el conocimiento de los métodos a emplear, y las metodologías que pueden utilizarse para desarrollar la herramienta propuesta.
- Capítulo 2, Herramientas Usadas presenta el software y algoritmos que se emplearon para el desarrollo del sistema propuesto.
- Capítulo 3, Estado del Arte muestra algunos trabajos relacionados al que se pretende desarrollar en esta investigación y finaliza con una comparación entre ellos y el propuesto.
- Capítulo 4, Metodología de la Investigación presenta la planeación de las etapas necesarias para cumplir el objetivo general del trabajo.
- Capítulo 5, Planeación de la Construcción de la Herramienta detalla los pasos que se siguieron para diseñar el sistema.
- Capítulo 6, Desarrollo de la Herramienta Propuesta ilustra el conjunto de actividades que se realizaron para construir los componentes del sistema.

-
- Capítulo 7, Validación de la Herramienta describe los instrumentos de medición usados por los especialistas que manejaron el sistema para evaluarlo y los resultados que se obtuvieron.
 - Conclusiones contiene la resolución de la hipótesis expuesta en este trabajo.

Marco Teórico

Para comenzar a comprender el fundamento de este trabajo, lo primero que se debe hacer es entender cuál es la naturaleza del problema y la de las herramientas que se van a usar con la intención de contribuir a su solución.

A continuación, se abordan temas clave en la elaboración de esta tesis y se dan explicaciones detalladas de cada uno de ellos con el fin de contextualizar a quien esté leyendo.

1.1. Suicidio

Existen varias definiciones del suicidio como la dada por la OMS (5) que se mencionó previamente: Lesión autoinfligida con intención de causar la muerte de la persona.

Así también está la propuesta por Shneidman (6), quien considera al suicidio como un acto consciente de aniquilación autoinducida, mejor entendida como un malestar multidimensional en las necesidades de un individuo que representa un problema para el cual el suicidio se percibe como la mejor solución.

En ambas definiciones, hay dos similitudes importantes, que son el daño autoinfligido y la intención de morir. Además, de la última se puede rescatar también el hecho de que se trata de una decisión que toma el individuo con base en cúmulo de problemas de diferentes orígenes.

1.1.1. Ideación suicida

Barrientos de Llanos et al (7) define a la ideación suicida como el deseo de morir, la idea de la propia muerte por alguna causa, la idea de matarse, pero sin otra elaboración, la misma idea, pero con planes inespecíficos de cómo hacerlo, así como la planificación detallada de la conducta a seguir, pero sin haber realizado alguna acción.

Este concepto se desarrolla en los adultos mayores de forma diferente a los jóvenes y es ciertamente más difícil de detectar debido a que las personas de la tercera edad reflejan menos señales de aviso (8). La ideación suicida puede ser alimentada o disuadida respectivamente por ciertas condiciones conocidas como factores de riesgo y factores protectores, los cuales se explican en las secciones a continuación.

1.1.2. Factores de riesgo

A pesar de que el suicidio es una causa de muerte prevenible, existe una gran dificultad en evitar que una persona se autolesione con la intención de morir. Esto es debido a que no hay parámetros concretos que indiquen que alguien está en riesgo o que, si lo está, se pueda asegurar la integridad de su vida. Es un asunto complejo porque se trata de una consecuencia de múltiples causas. Aunque se pudiera pensar que ocurre por una sola razón, en realidad se trata de la suma de una serie de factores de riesgo y precipitantes que van acumulándose hasta tener más influencia en que una persona abrace la idea de finalmente quitarse la vida y por lo tanto, ningún factor explica por sí solo un suicidio. Los factores de riesgo, como se sugirió anteriormente, son los motivos y eventos que ocasionan y producen un dolor agobiante que perjudica la calidad de vida de un individuo, y hacen que la probabilidad de que una persona se suicide sea más alta.

El Gobierno de Navarra en España (9), la Organización Panamericana de la Salud (OPS)(10), la OMS (11), entre otros autores (12) (13), han clasificado diferentes factores

de riesgo, de los cuales, se ha hecho una compilación que se presenta a continuación:

1.1.2.1. Factores sociales

Se conoce como factores sociales a las influencias externas que un individuo recibe de su propio entorno e impactan en lo que piensa y siente, por ejemplo:

- Edad: Con la edad aumenta el número de suicidios.
- Genero: Mientras que los suicidios completados son más comunes entre los hombres, las mujeres tienen un mayor riesgo en las otras conductas suicidas.
- Profesión: Las profesiones con mayor incidencia de suicidio son las de ámbito sanitario, ámbito educativo y de fuerzas de seguridad.
- Crisis social, económica y/o sanitaria.
- Suicidios en el entorno / suicidios mediáticos con ausencia o inadecuada información: Sensacionalismo, simplificación de las causas, divulgación de métodos utilizados.
- Inadecuado o ausencia de acompañamiento- apoyo social: Soledad objetiva, soledad sentida.
- Falta de recursos de ayuda accesibles y eficientes. Abusos sexuales, maltrato, negligencia (también en personas mayores).
- Contextos de vulnerabilidad social y/o marginación social:
 - LGTBI: se estima que el 30 % de los suicidios en jóvenes están relacionados con este factor.
 - Acoso-bullying en todas sus variantes, humillaciones (sobre todo en adolescentes).

1. MARCO TEÓRICO

- Violencia de género
- Población migrante
- Personas sin hogar
- Personas ingresadas en prisión
- Fácil acceso a métodos de suicidio

1.1.2.2. Factores personales

Se trata del conjunto de elementos inherentes y propios de los individuos, que se van consolidando al perfil de la personalidad. Abordándolos como factores que contribuyen en la decisión del suicidio. Algunos de estos son:

- Intentos de suicidio previos: Se dice, es el factor de riesgo más importante, pues predice entre un 50-80 % de los casos. Una de cada 3 personas repetirá un intento de suicidio. Entre el 10 %-20 % lo repetirán en el primer año. Es muy importante extremar el cuidado y la atención en los meses posteriores al intento, incluso hasta un año. Entre el 10-15 % de personas que lo han intentado posteriormente lo consuman.
- Ideación de suicidio: El 26 % de personas con ideación realizarán una tentativa sin tener una planificación.
- Planificación de suicidio: El 70 % de personas con un plan de suicidio realizará una tentativa. No siempre que se da un intento de suicidio hay necesariamente un plan de cómo hacerlo.
- Presencia de autolesiones: Las autolesiones son un factor predictor de intentos de suicidio en adolescentes y jóvenes adultos.
- Enfermedades físicas: Más riesgo en pacientes con enfermedades crónicas que provocan dolor físico recurrente, discapacidad, limitación funcional y pronóstico

grave. En el 25 % de los suicidios está presente este factor. En personas mayores llega al 80 %.

- Biografía de sufrimiento, diversas causas se asocian a este factor, como:
 - Sensación de ser una carga para los demás y de que estarán mejor sin él/ella.
 - Sensación de atrapamiento.
 - Sentimiento de fracaso.
 - Falta de sentido de vida.
 - Desbordamiento emocional.

1.1.2.3. Factores psicológicos

Estos se refieren a los estándares conductuales, cognitivos y emocionales que existen en la calidad de vida de las personas, reflejados en su estado y prácticas de salud. En este contexto se observa el listado siguiente:

- Desesperanza: Es el factor psicológico más influyente en el riesgo de la conducta suicida. Pérdida de la esperanza e ilusión para afrontar los problemas.
- Baja autoestima: No sentirse bien con las características personales que uno tiene no aceptándose a sí mismo. Tiene un alto componente comparativo con otras personas, grupos sociales y genera desequilibrio emocional.
- Rigidez cognitiva: Dificultad e incapacidad de flexibilizar su visión y manejar otras alternativas. Es la visión “en túnel”.
- Falta de habilidades sociales: Dificultades de comunicación, de resolución de problemas. Falta de asertividad.
- Impulsividad, irritabilidad: Sobre todo, en jóvenes. No siempre hay una elaboración premeditada del suicidio.

1.1.2.4. Factores precipitantes

Como el nombre lo sugiere, este tipo de factores agudizan la inestabilidad emocional y aumentan el sufrimiento, generando un desequilibrio en la salud mental que conllevan un mayor riesgo de suicidio y que hacen precisa una mayor atención. Pueden incluir todas las situaciones de crisis, de estrés y de pérdidas en todos los ámbitos que conforman el entorno de una persona y según la edad del individuo se clasifican de la siguiente manera:

- Comunes a todas las edades: enfermedades incapacitantes o dolorosas, pérdida de seres queridos, conflictos interpersonales, existencia de un entorno suicida, experiencias traumáticas (agresiones, malos tratos, acoso, abuso, desastres naturales), consumo o aumento de tóxicos, condenas judiciales, cambios vitales bruscos no deseados.
- Específicos de edades cortas entre la infancia y la adolescencia: situaciones de vulnerabilidad relacionadas con la identidad sexual o con la orientación sexual, fracaso escolar, cyberbullying, bullying, humillaciones, abuso sexual, violación, maltrato físico, exposición a violencia, separación o pérdida de amigos, cambios de residencia o de entorno social, desengaño amoroso, muerte de un ser querido, problemas disciplinarios escolares, presión escolar, entorno suicida, embarazo prematuro, separación de los padres, conflictos conyugales.
- Específicos de la edad madura y vejez: fracaso profesional, pérdida de empleo y dificultades económicas, pérdida de estatus social, ruptura conyugal o desengaño amoroso, jubilación, empeoramiento o debut de enfermedades, pérdida de capacidad funcional, soledad, institucionalización, cambio periódico de domicilio, malos tratos, viudedad, pérdida de interés por la vida, pérdida de la fe.

1.1.2.5. Trastornos mentales, como factor de riesgo

La mayoría de las personas que presentan un trastorno mental no se suicidan. Sin embargo, los trastornos mentales llegan a relacionarse estrechamente con el suicidio (14) y algunos de ellos se muestran a continuación:

- Trastornos mentales sin diagnosticar: Por lo tanto, no cuentan con soporte de ayuda médica farmacológica ni terapéutica.
- Depresión: Presenta un riesgo 20 veces superior a la población general. De las personas diagnosticadas con depresión, el 15 % mueren por suicidio. Las tentativas del suicidio pueden llegar a un 40 %. Se asume que más del 65 % de los suicidios e intentos de suicidio se relacionan con algún grado de depresión.
- Trastorno bipolar: Presenta un riesgo 15 veces superior a la población general.
- Esquizofrenia: Presenta un riesgo 9 veces superior a la población general.
- Trastorno límite de personalidad: Presenta un riesgo 4-8 veces superior a la población general.
- Trastorno de la conducta alimentaria: Presenta un riesgo 4 veces superior a la población general.
- Abuso de alcohol y otras sustancias: No es sólo un factor de riesgo sino también un factor precipitante. El riesgo es 6 veces mayor que en la población general.

1.1.3. Factores protectores

Los factores protectores, a diferencia de los de riesgo, son aquellos que se refieren a las condiciones que reducen las posibilidades de que una persona cometa autolesiones con intención fatal al aumentar su resistencia a los trastornos que conducen a tal

1. MARCO TEÓRICO

decisión. La OPS (10), clasifica estos factores en cuatro diferentes tipos, los cuales se abordan en las siguientes secciones.

1.1.3.1. Individuales

Este tipo de factores se refieren a aquellos que están relacionados exclusivamente con la actitud de un individuo por cuidar de si mismo. Tales como:

- Buen estado de salud
- Hábitos saludables
- Adecuada nutrición
- Actitud y carácter positivos
- Buena relación con los pares
- Apropriadas aptitudes sociales
- Sentido de esperanza y optimismo
- Autoestima adecuada
- Manejo del estrés
- Adaptabilidad

1.1.3.2. Familiares

Los factores familiares están más relacionados con la relación que tiene el individuo con su propia familia y cómo estos han impactado desde su origen hasta la actualidad. Algunos ejemplos pueden ser:

- Satisfacción de las necesidades básicas aseguradas mediante el empleo estable y digno

- Estabilidad y coherencia del núcleo familiar
- Estímulo y refuerzo de actividades y comportamientos positivos
- Vida regular en el colectivo
- Reconocimiento de logros
- Apoyo a metas positivas
- Promoción de la amistad, la solidaridad, la tolerancia y la inclusión
- Establecimiento de límites sociales apropiados de conducta

1.1.3.3. Comunitarios

Los factores comunitarios son aquellos que tienen que ver con la comunidad en la que vive una persona. Se relacionan a actividades con las cuales la sociedad en general contribuye al sano desarrollo de sus miembros.

Algunos de ellos pueden ser:

- Oportunidades educativas y profesionales
- Actividades para niños y adolescentes (deportes y actividades educativas)
- Apoyo social
- Apoyo a las familias con necesidades especiales
- Entorno sin riesgos (ausencia de violencia, uso de drogas o contaminación ambiental)
- Construcción de espacios públicos seguros (barrios, parques, plazas, edificios, caminos)
- Oportunidades para el desarrollo económico sostenido (capacitación y créditos para iniciar pequeñas empresas destinados a personas en situación de pobreza)

1.1.3.4. Institucionales

Estos factores, si bien son parecidos a los anteriormente mencionados (comunitarios), se refieren más a aquellas condiciones que la sociedad y el gobierno han fomentado a través de instituciones privadas o públicas para el bienestar de la población en general. Estas instituciones pueden dedicarse a diversos sectores, algunas de ellas pueden ser las siguientes:

- Organizaciones basadas en las necesidades de las personas y que ofrecen buen trato
- Instituciones proactivas en la defensa del medio ambiente
- Instituciones facilitadoras de la atención a personas con desventajas (capacidades diferentes, minorías étnicas) y promotoras de la responsabilidad social por la salud mental de la población.

1.1.4. Panorama actual del suicidio en México

Según el INEGI (1), en 2021, las estadísticas de mortalidad reportaron que del total de fallecimientos en el país (1,093,210), 8,351 fueron por lesiones autoinfligidas. Esto representa una tasa de suicidio de 6.5 por cada 100 mil habitantes. Como se puede apreciar en la Figura 1.1, de los decesos por esta causa, destaca que los hombres tienen una tasa de 10.9 suicidios por cada 100 mil (6,785). La tasa para las mujeres es menor: 2.4 por cada 100 mil (1,552).

En la Figura 1.2, se observa como el suicidio en personas de 15 a 29 aumentó en los últimos años. En 2015 se estimó una tasa de 8.1 muertes por lesiones autoinfligidas por cada 100 mil personas de las edades mencionadas. En 2021, la tasa de suicidios para el mismo grupo fue de 10.4 por cada 100 mil personas. En el caso de los hombres de 15 a 29 años, el aumento en el riesgo de suicidio aumentó de 12.4 en 2015 a 16.2 en 2021

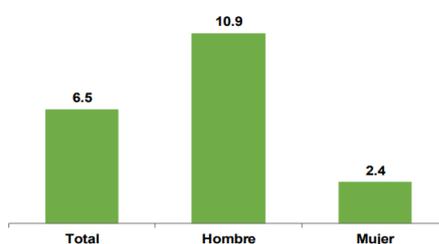


Figura 1.1: Tasa de suicidios según sexo (Por cada 100 mil habitantes) Fuente: INEGI (2022)

por cada 100 mil en esas edades. Sin embargo, no es el único sector de la población que requiere de atención en este aspecto, ya que los adultos mayores también se encuentran entre las poblaciones más vulnerables a concebir la idealización suicida.

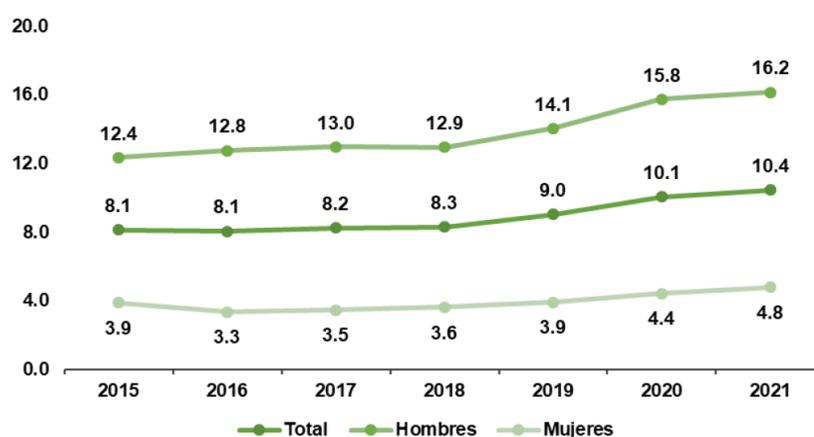


Figura 1.2: Tasa de suicidio en la población de 15 a 29 años desde 2015 hasta 2021 (Por cada 100 mil habitantes) (1)

Como se puede ver en la Figura 1.3 proporcionada por INEGI, los adultos mayores de 60 años tienden a pensar en suicidarse. Tomando en cuenta que la consumación del suicidio está ligada con trastornos psicológicos como la depresión y los adultos mayores suelen padecerla por diversos factores (Los cuales se abordaron en secciones anteriores), es de esperar que en el futuro haya un aumento en la cantidad de muerte autoinducida en esta parte de población.

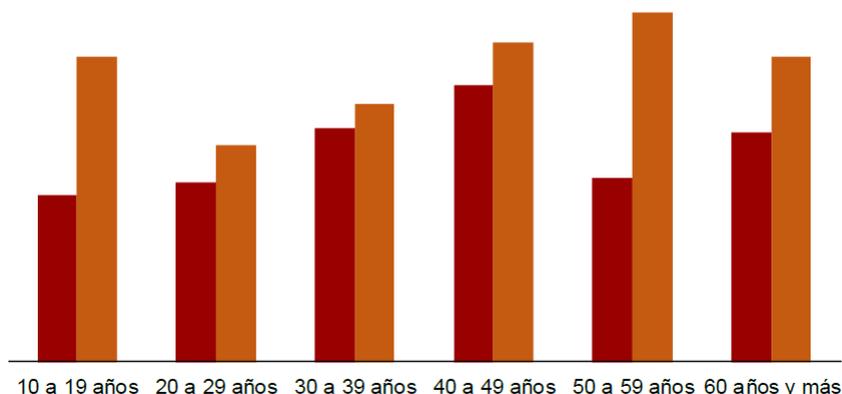


Figura 1.3: Porcentaje de la población de 10 años y más que ha pensado suicidarse, por sexo y grupo de edad (2018) ⁽¹⁾

1.2. Inteligencia Artificial

Rich et al (15) definen a la Inteligencia Artificial (IA) como el estudio de cómo lograr que las computadoras realicen tareas que, por el momento, los humanos hacen mejor.

Desarrollando más a fondo la idea anterior, la IA intenta emular la inteligencia humana al realizar actividades que están vinculadas con el pensamiento humano como la toma de decisiones, resolución de problemas, aprendizaje, entre otras. Esto lo hace mediante modelos computacionales que tienen su base en los cálculos matemáticos que hacen posible que una entidad informática sea considerada como inteligente.

Una vez aclarado el término de IA, se puede ver que en la actualidad, esta se manifiesta de varias formas en la vida diaria. Algunos ejemplos son:

- Los chatbots utilizan la IA para comprender más rápido los problemas de los clientes y proporcionar respuestas más eficientes.
- Los asistentes inteligentes utilizan la IA para analizar información crítica proveniente de grandes conjuntos de datos de texto libre para mejorar la programación.

- Los motores de recomendación pueden proporcionar recomendaciones automatizadas para programas de TV según los hábitos de visualización de los usuarios.

Desde el punto de vista corporativo de una empresa como Oracle (16), la IA trata mucho más sobre el proceso y la capacidad de pensamiento super poderoso y el análisis de datos que sobre cualquier formato o función en particular. Aunque la IA muestra imágenes de robots de aspecto humano de alto funcionamiento que se apoderan del mundo, la IA no pretende reemplazar a los humanos. Su objetivo es mejorar significativamente las capacidades y contribuciones humanas. Eso la convierte en un activo empresarial muy valioso.

1.2.1. Inteligencia Computacional

Se le conoce como Inteligencia Computacional (IC) al estudio del diseño de entidades con la capacidad de comunicarse con un entorno, procesar las señales que recibe de él y, responder o actuar en su entorno de una forma previamente establecida y mejorando con el tiempo. Estas entidades también son conocidas como agentes inteligentes (17).

La IC se ocupa de la teoría, diseño, desarrollo y aplicaciones de paradigmas computacionales motivados lingüística y biológicamente, poniendo énfasis en las redes neuronales, algoritmos genéticos, programación evolutiva, sistemas difusos y sistemas inteligentes híbridos.

Estudia problemas para los que no hay algoritmos efectivos, bien porque no es posible formularlos o porque requieren unos tiempos de ejecución exponenciales (no polinomiales).

Una buena parte de la investigación se centra en funciones cognitivas de bajo nivel, percepción, reconocimiento de objetos, análisis de señales, descubrimiento de estructuras en conjuntos de datos, asociación y control. Los métodos para resolver este tipo de problemas incluyen aprendizaje supervisado y no supervisado mediante sistemas adaptativos, y compaginan, no sólo los enfoques neuronales, difusos o evolutivos, sino

también los enfoques probabilísticos y estadísticos, como las redes Bayesianas y los métodos basados en núcleos.

Las funciones cognitivas de alto nivel se requieren para resolver problemas que no se pueden resolver mediante estos algoritmos y desarrollan razonamiento, representación compleja del conocimiento, memoria episódica, planificación o comprensión del razonamiento simbólico.

La IC es una metodología para el desarrollo de computación que muestra capacidad para aprender y afrontar nuevas situaciones, de manera que el sistema presenta algunos atributos del razonamiento tales como la generalización, descubrimiento, asociación y abstracción (18).

1.2.2. Reconocimiento de patrones

El término de reconocimiento de patrones se refiere a un procesamiento de información que tiene una gran importancia práctica que da solución a un amplio rango de problemas. Algunos de estos problemas son resueltos por los humanos sin mucho esfuerzo. Sin embargo, en muchos casos, la solución a estos problemas, usando computadoras, se vuelve extremadamente dificultosa (19). Reducir estas dificultades, construyendo métodos y algoritmos, es el objetivo principal del Reconocimiento de patrones como una disciplina de la Ingeniería de Sistemas.

En términos más concretos, el Reconocimiento de Patrones como una disciplina científica tiene el objetivo de clasificar objetos en un número específico de categorías o clases. Dependiendo de la aplicación, estos objetos pueden ser imágenes, sonidos, olores, en general, señales producto de mediciones que deben ser clasificadas. Estos objetos se denotan con el término genérico de patrones.

La producción industrial se está automatizando y esto ha logrado hacer del Reconocimiento de Patrones un tema de investigación importante en el campo de la Ingeniería de Sistemas, como parte integral de los Sistemas Inteligentes (20).

1.2.3. Minería de Datos

La Minería de Datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

Estos patrones y tendencias se pueden recopilar y definir como un modelo de Minería de Datos. Estos modelos se pueden aplicar en escenarios como los siguientes:

- Pronóstico: cálculo de las ventas y predicción de las cargas del servidor o del tiempo de inactividad del servidor.
- Riesgo y probabilidad: elección de los mejores clientes para la distribución de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- Recomendaciones: determinación de los productos que se pueden vender juntos y generación de recomendaciones.
- Búsqueda de secuencias: análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos.
- Agrupación: distribución de clientes o eventos en grupos de elementos relacionados, y análisis y predicción de afinidades.

Según Microsoft (2), la generación de un modelo de Minería de Datos forma parte de un proceso mayor que incluye desde la formulación de preguntas acerca de los datos y la creación de un modelo para responderlas, hasta la implementación del modelo en un entorno de trabajo. Este proceso se puede definir mediante los pasos siguientes:

- Definición del problema

- Preparación de datos
- Explorar los datos
- Creación de modelos
- Exploración y validación de modelos
- Implementar y actualizar los modelos

En la Figura 1.4 se describen las relaciones entre cada paso del proceso que apoya el procesamiento de Datos (2).

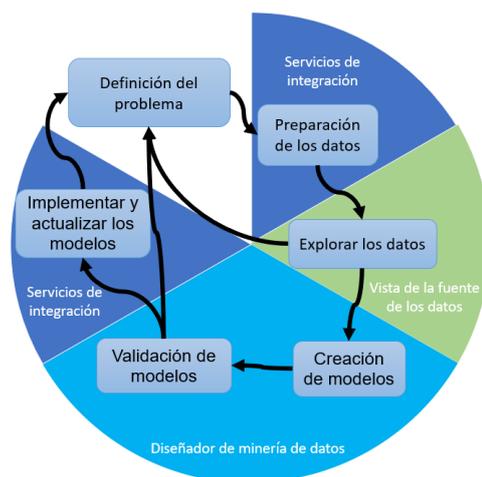


Figura 1.4: Modelo de Minería de Datos (Elaboración propia basada en (2))

El proceso que se ilustra en el diagrama es cíclico, lo que significa que la creación de un modelo de Minería de Datos es un proceso dinámico e iterativo. Una vez que ha explorado los datos, puede que descubra que resultan insuficientes para crear los modelos de Minería de Datos adecuados y que, por tanto, debe buscar más datos. O bien, puede generar varios modelos y descubrir entonces que no responden adecuadamente al problema planteado cuando los definió y que, por tanto, debe volver a definir el problema. Es posible que deba actualizar los modelos una vez implementados debido

a que haya más datos disponibles. Puede que haya que repetir cada paso del proceso muchas veces para crear un modelo adecuado (2).

Existen varias funcionalidades de la minería de datos. Entre ellas se encuentran la caracterización y la discriminación; la extracción de patrones frecuentes, asociaciones y correlaciones; la clasificación y la regresión; el análisis de conglomerados; y el análisis de valores atípicos. Las funcionalidades de la minería de datos se utilizan para especificar los tipos de patrones que deben encontrarse en las tareas de minería de datos. En general, estas tareas pueden clasificarse en dos categorías: descriptivas y predictivas. La minería descriptiva tiene por objetivo caracterizar las propiedades de los datos de un conjunto de datos objetivo. Mientras que la minería predictiva realiza inducciones sobre los datos actuales para hacer predicciones.

1.2.4. Aprendizaje Automático

El termino de Aprendizaje Automático (AA) no es nuevo dentro de las ciencias computacionales, aunque al igual que las otras ramas de la Inteligencia Artificial vistas anteriormente, también ha descubierto su auge en la actualidad, gracias a los avances tecnológicos que han dotado a las computadoras con capacidades superiores a las de la época en la que esos conceptos se dieron a conocer.

Arthur Samuel (21), pionero en el AA, lo define como el campo de estudio que da a los ordenadores la capacidad de aprender sin ser programados específicamente.

De una forma más técnica y dirigida a la parte de la ingeniería, Mitchell (22) dice que un programa informático aprende de la experiencia E con respecto a alguna tarea T y alguna medida de rendimiento R , si su rendimiento en T , medido por R , mejora con la experiencia E .

En esencia, vemos que el propósito del AA es construir un programa informático que pueda ser alimentado o entrenado con un conjunto de datos con el cual sea capaz proporcionar información para llevar a cabo una determinada tarea, evitando en lo

posible la intervención del ser humano y tomando en cuenta el costo computacional.

Al conjunto de datos ya mencionado, se le puede llamar conjunto de entrenamiento. Este va a estar compuesto de una serie de ejemplos que van a conformar la parte de la experiencia en y como ya vimos, es fundamental para que el AA sea posible. A cada uno de estos ejemplos se le conoce como muestra o instancia de entrenamiento.

Es necesario aplicar el AA en una entidad artificial para que pueda ser considerada inteligente. Si una máquina no es capaz de aprender cosas nuevas, eso imposibilitaría su capacidad de adaptarse al medio, condición que es indispensable para cualquier ser dotado de inteligencia.

Entre las aplicaciones que se pueden observar en el día a día y han mejorado la calidad de vida de las personas, se tienen, por ejemplo:

- Algoritmos de recomendaciones.
- Análisis de imágenes y detección de objetos.
- Detección de fraude.
- Sistemas de atención al cliente o Chatbots.

Hay diferentes formas de clasificar a los modelos de aprendizaje dentro del AA, pero en la que está basa la cantidad y el tipo de supervisión que reciben durante el entrenamiento, se aprecia de una forma más clara como es que estos modelos están estrechamente relacionados con otras ramas de la IA. Se explican en los apartados a continuación.

1.3. Tipos de aprendizaje

Como ya se dijo anteriormente, la Inteligencia Artificial está dividida en diferentes ramas como las que se tomaron en cuenta en los apartados anteriores. Estas tienen

diferentes técnicas de aplicación y poseen diferentes objetos de estudio, pero también tienen varios aspectos en común y es por ello que tienen la posibilidad de fusionarse y complementarse para desarrollar sistemas que aprendan de conjuntos de datos y aporten algún beneficio a aquellos para quienes fueron diseñados.

Existen diferentes clases de aprendizajes para las computadoras y cada uno de ellos posee recursos para implementarse. Algunos de ellos se abordarán con mayor detalle en las siguientes secciones.

1.3.1. Aprendizaje supervisado

En el aprendizaje supervisado, el conjunto de datos usado para el entrenamiento (conjunto de entrenamiento) que se introduce en el algoritmo, incluye las soluciones deseadas, que pueden llamarse etiquetas o clases. Es decir, se le dice previamente al modelo que es lo que se quiere que aprenda.

Una tarea típica de aprendizaje supervisado es la clasificación. Esta consiste dar por salida una etiqueta discreta, es decir, una etiqueta dentro de un conjunto finito de valores posibles. Del mismo modo, los modelos pueden dar como resultado una clasificación binaria si tienen que predecir entre dos etiquetas (clases) o múltiples etiquetas (multiclase), cuando son más de dos.

El filtro de spam es un buen ejemplo de ello: se entrena con muchos correos electrónicos de ejemplo junto con su clase (spam o ham), y debe aprender a clasificar los nuevos correos. En la Figura 1.5 se ve este ejemplo ilustrado.

Otra tarea típica, se denomina regresión y consiste en predecir un valor numérico objetivo, a partir de un conjunto de características denominadas predictores. Para entrenar el sistema, es necesario brindarle un conjunto de entrenamiento, valga la redundancia, que contendrá tanto etiquetas como predictores.

Supóngase un caso en el que se tomara como etiqueta el precio de una computadora, obteniéndolo a partir de sus predictores, que serían características como su

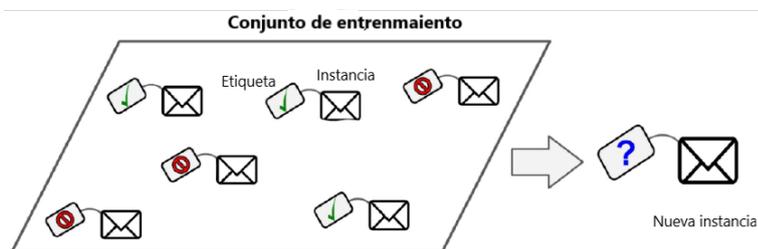


Figura 1.5: Ejemplo de un conjunto de entrenamiento etiquetado para aprendizaje supervisado (filtro de spam)⁽³⁾

procesador, marca, almacenamiento, etc. Para entrenar un modelo de regresión en este caso, se debería entrenar con muchos ejemplos de computadoras con sus precios y las características mencionadas en un inicio.

No está de más mencionar que algunos modelos de clasificación pueden utilizarse también para regresión, y viceversa. Geron ⁽³⁾ indica que la regresión logística se utiliza habitualmente para la clasificación, esto debido a que puede dar como resultado un valor que corresponde a la probabilidad de pertenecer a una clase determinada. Retomando el ejemplo del filtro de spam, habría un 20 % de probabilidades de ser spam.

Por último, se hace mención de algunos de los algoritmos de aprendizaje supervisado más importantes que se encontraron en la literatura:

- K vecinos más cercanos
- Regresión lineal
- Regresión logística
- Máquinas de vectores soporte
- Árboles de decisión y random forest
- Redes neuronales

1.3.2. Aprendizaje no supervisado

A diferencia del aprendizaje supervisado, el aprendizaje no supervisado trabaja con datos que no han sido etiquetados, por lo que no hay una etiqueta que predecir. En términos más generales, el modelo intenta aprender sin una guía.

Estos modelos se usan principalmente para resolver problemas donde es necesario analizar los datos con la intención de extraer nuevo conocimiento o agrupar entidades por afinidad.

La agrupación consiste en encontrar diferentes grupos dentro de los elementos de los datos. Los modelos de agrupamiento se encargan de encontrar una estructura en los datos de tal forma que las instancias del mismo grupo sean más similares entre sí que con los de otros grupos diferentes.

Tomando como punto de partida el modelo anterior, los modelos de visualización también son buenos ejemplos de algoritmos de aprendizaje no supervisado y se pueden usar para complementar a los algoritmos de agrupación. Ya que se les entrena con una gran cantidad de datos sin clasificar y dan por resultado una representación 2D o 3D de los datos que facilita la comprensión de la información obtenida para quien esté a cargo de analizarla. Así también, los modelos de visualización intentan salvaguardar la estructura tanto como les es posible para lograr un mejor entendimiento de cómo están organizados los datos y e incluso encontrar patrones que no se habían descubierto con anterioridad.

Otra de sus aplicaciones es reducir dimensionalidad o simplificar los conjuntos de entrenamiento. Lo que quiere decir que tiene por objeto reducir los datos sin perder demasiada información. Una forma de hacerlo es unificar varias etiquetas correlacionadas en una sola. Por ejemplo, el año de fabricación de una computadora puede estar fuertemente ligado con su sistema operativo, entonces, algoritmo de reducción de la dimensionalidad los combinará en una sola característica que represente que tan moderna

es la computadora. A lo anterior se denomina extracción de características.

También existen los modelos que se centran en la detección de anomalías, que no son otra cosa más que valores atípicos de un conjunto de datos. Uno de los usos más comunes de la detección de anomalías, es previo a entrenar otro modelo de aprendizaje. Durante el entrenamiento, el algoritmo recibe una mayor cantidad de instancias normales, por lo que aprende a distinguirlas y cuando ve una nueva instancia, puede decir si es común o anormal.

Al final, se encuentra el aprendizaje de reglas de asociación, cuyo propósito es buscar relaciones significativas entre las clases de conjuntos de datos de gran tamaño. Lo anterior hace que esta aplicación sea muy útil para resolver problemas de la vida real. Por último, se mencionen algunos de los algoritmos que forman parte de esta clasificación:

- Agrupación
 - K-Means
 - DBSCAN
 - Análisis jerárquico de conglomerados
- Detección de anomalías y novedades
 - Máquinas de vectores soporte de una clase
 - Bosque aislado
- Visualización y reducción de la dimensionalidad
 - Análisis de componentes principales
 - Núcleo PCA
 - Incrustación lineal local (LLE)
 - Incrustación estocástica de vecinos distribuida en t (t-SNE)

- Aprendizaje de reglas de asociación
 - Apriori
 - Eclat

1.3.3. Aprendizaje semisupervisado

El aprendizaje semisupervisado es la unión entre los dos tipos de aprendizaje descritos en los apartados anteriores. Existen algunos modelos que pueden trabajar con datos de entrenamiento parcialmente etiquetados, según la literatura consultada, es muy común ver que la mayoría de los datos de un conjunto, están sin etiquetar y, por lo tanto, la minoría restante son los que están etiquetados. Esto es a causa de que el proceso de etiquetado llega a ser económicamente costoso o implicar dificultades técnicas.

Se podría pensar que una solución práctica para el problema del etiquetado sería simplemente realizar el entrenamiento pasando por alto las instancias no etiquetadas o viceversa, sin embargo, estas pueden proporcionar información útil para mejorar el aprendizaje de un sistema que esté implementando un modelo de Aprendizaje Automático y los resultados que se obtengan para su posterior análisis.

Así pues, este tipo de aprendizaje se puede en dos tipos (23), según el objetivo del análisis que se busque hacer a el conjunto de datos.

Para comenzar, está el tipo introducido por Vapnik (24), al que también se puede llamar aprendizaje transducido. Este consiste en separar en dos partes el conjunto de datos, en uno se van a guardar las instancias que, si están etiquetadas y el resto, que no está etiquetado, se reserva en el otro conjunto. Esto se hace con el propósito de intentar predecir los valores de la clase en el conjunto de instancias que falta etiquetar.

Al otro se le conoce como aprendizaje inductivo y se centra en obtener una función que anticipe de lo que ocurrirá de acuerdo al análisis de las características del conjunto

de datos, usando todo el conjunto, es decir, tanto los datos que han sido etiquetados como los que no.

Para finalizar este apartado, se mencionan dos de los algoritmos más comunes y útiles en conjuntos de datos propios del aprendizaje semisupervisado:

- Self-training
- Co-training

1.4. Metodologías

Lara (25) define a la metodología como un conjunto de aspectos operativos que se tienen en cuenta para cumplir los objetivos un proyecto. Por lo anterior, se trata de un recurso esencial que permite realizar proyectos de una forma organizada, estructurada y estandarizada.

Es por ello que definir una metodología es muy importante si se desea tener una mejor organización del tiempo, los recursos, las herramientas disponibles, aumentar la eficiencia en el trabajo y reducir la probabilidad de que existan errores o percances para llegar a la culminación de un proyecto (26).

Existen diversas metodologías de trabajo que se pueden adoptar, dependiendo de que tanto se adapten a las metas que se pretenden alcanzar en el proyecto en particular. Se dividen en dos categorías, tradicionales y ágiles, las cuales se mencionan en los siguientes apartados.

1.4.1. Tradicionales

Las metodologías de tipo tradicional se han empleado desde el año 1970 con gran éxito para el desarrollo de proyectos. Consisten en hacer que un proyecto pase por una serie de ciclos de vida en forma secuencial, donde típicamente se incluyen las fases de inicio, planificación, ejecución, seguimiento y cierre (27).

Algunos ejemplos de metodologías tradicionales se mencionan a continuación.

1.4.1.1. Cascada

La metodología más representativa de la clasificación tradicional fue propuesta por Winston Royce (28), la cual es denominada “Cascada” por la posición de las fases que la componen, puesto que estas van avanzando de una forma que asemeja a la caída por una cascada, como se puede apreciar en la Figura 1.6. Se trata de un proceso rigurosamente ordenado, donde para iniciar una etapa, la predecesora debe haber sido concluida.

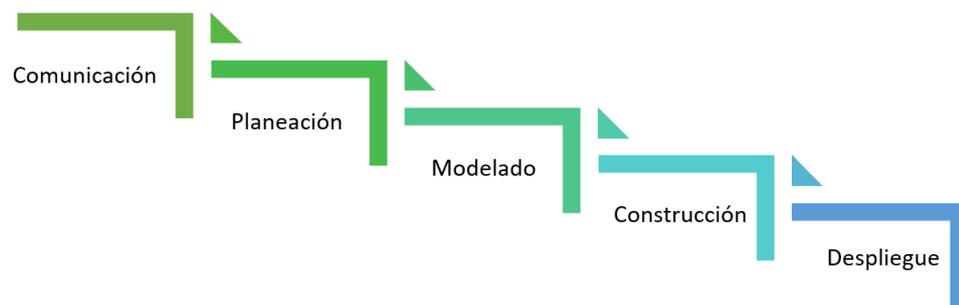


Figura 1.6: Metodología de Cascada (Elaboración propia)

Es conveniente el uso de esta metodología cuando el proyecto a realizar está relacionado con el desarrollo de sistemas computacionales que llevan a cabo transacciones que permiten la gestión de información del exterior. También es necesario resaltar el hecho de que es una metodología cuya eficiencia es superior cuando se trata de organizar grandes presupuestos y equipos de trabajo que colaboran en el desarrollo del proyecto (4).

El computólogo que propuso esta metodología refiere que sus beneficios son más notorios cuando no hay fechas de entrega rigurosas, ya que los desarrolladores cuentan con suficiente tiempo para implementar cada fase. Cabe resaltar que, en esta metodología, los inconvenientes o riesgos disminuyen si los requerimientos son claros y están establecidos en la primera fase del proyecto.

- **Comunicación:** Es el inicio del proyecto, durante esta etapa, se hacen varias reuniones entre los desarrolladores y el usuario/cliente para recabar los requerimientos de dicho proyecto.
- **Planeación:** Se hace la estimación del tiempo, capital económico y recurso humano necesarios para el desarrollo del proyecto.
- **Modelado:** Esta etapa es aquella en la que se analiza lo realizado en las etapas anteriores para hacer un diseño que servirá como guía posteriormente para el desarrollo del proyecto.
- **Construcción:** Como el nombre lo sugiere, en esta parte de la metodología, se construye el proyecto atendiendo todas las necesidades previstas en etapas anteriores y se hacen las pruebas necesarias para comprobar su funcionamiento.
- **Despliegue:** El usuario recibe el proyecto terminado y también puede hacer una retroalimentación de cualquier índole y el equipo encargado del proyecto le brinda una asistencia para resolver cualquier duda o solicitud respecto a la versión obtenida del proyecto.

1.4.1.2. Espiral

Barry Boehm (29) presentó una metodología llamada “Espiral”, que implicó la primera modificación importante a los modelos de desarrollo de software. Sus fases están ilustradas en la Figura 1.7, en donde se puede apreciar.....

Tal modelo brinda la posibilidad de analizar detalladamente las etapas de desarrollo de un proyecto y al igual que en la metodología de Cascada, es necesario que los objetivos sean establecidos previamente a realizar cualquier acción que lleve a alcanzarlos. Una característica principal de esta metodología es que puede existir una retroalimentación a alguna de las fases anteriores a la que se está desarrollando, de tal manera que se mejore el proceso que se está siguiendo.

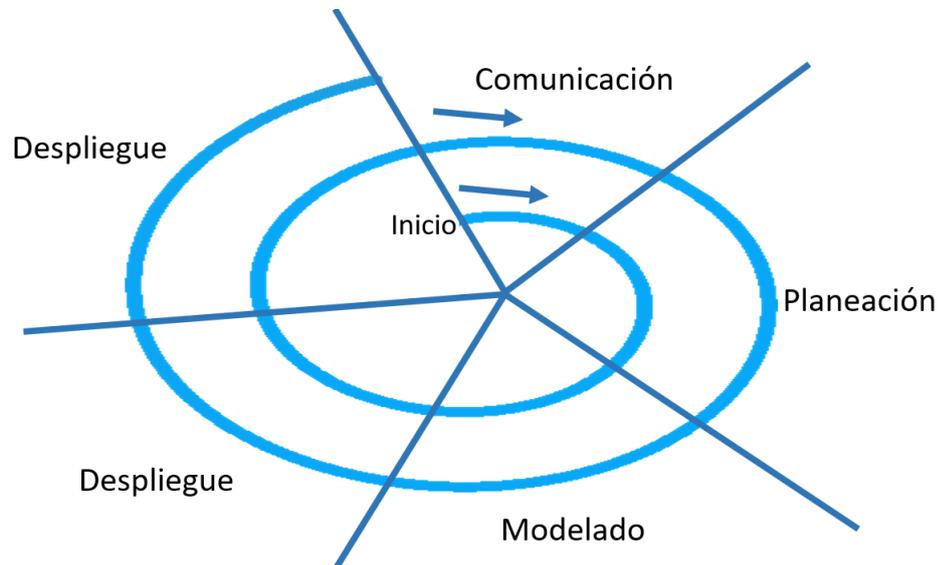


Figura 1.7: Metodología de Espiral (Elaboración propia basada en (4))

La anterior característica hace que esta metodología sea flexible, por lo que han habido diversas versiones modificadas que se adaptaron a las necesidades de diferentes equipos de desarrollo. Además, esta flexibilidad también es una característica valiosa al establecer plazos de entrega, debido a que algunas etapas pueden requerir de una prórroga para su satisfactoria culminación (30).

Por lo regular, las fases son las mismas que en la metodología de cascada, pero pueden variar dependiendo de la versión de cada autor.

1.4.1.3. Incremental

Esta metodología descompone un proyecto en una serie de secuencias lineales (llamadas incrementos) en forma escalonada. Esta descomposición se da a medida que avanza el calendario de actividades.

La metodología de desarrollo incremental se centra en que en cada incremento se entrega un producto que ya es capaz de funcionar, lo que permite ver resultados de una forma más rápida en comparación con la metodología de Cascada. Los primeros

1. MARCO TEÓRICO

incrementos son versiones desnudas del producto final, pero proporcionan una versión que sirve al usuario y también le dan una pauta de comparación entre ella y las que surgirán en las siguientes etapas (4). La forma en la que se comporta la metodología incremental se aprecia mejor en la Figura 1.8.

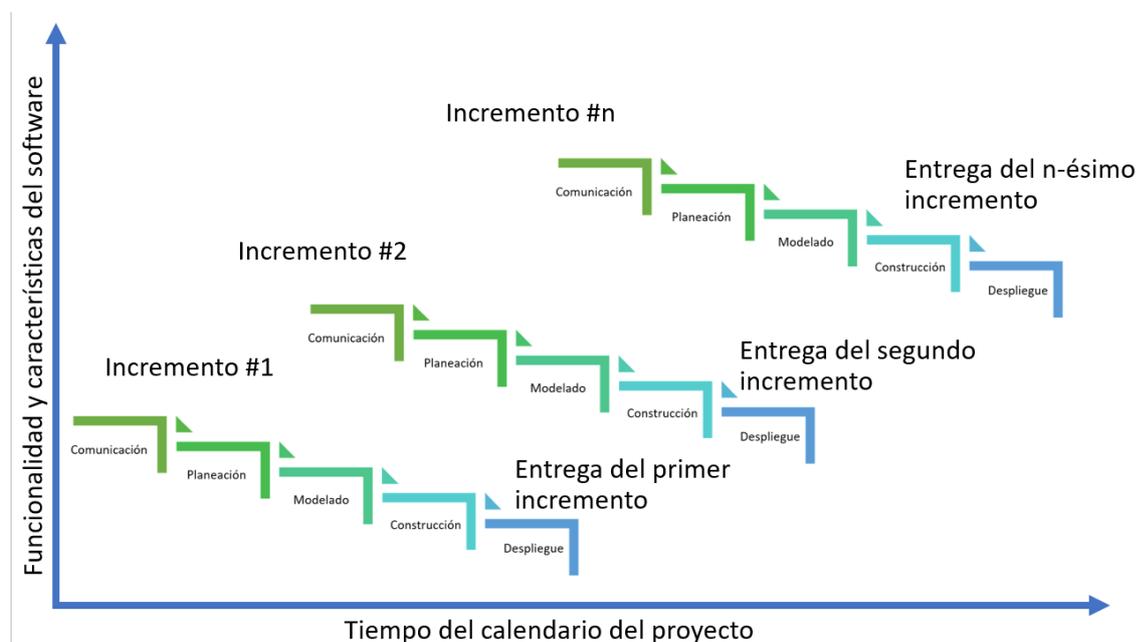


Figura 1.8: Metodología incremental (Elaboración propia)

Un hecho interesante acerca de esta metodología, es que es útil cuando los equipos de trabajo son reducidos. La idea es que el número de integrantes de los equipos sea pequeño en los primeros incrementos y vaya aumentando según sea necesario.

Al igual que espiral, las etapas de esta metodología son las mismas que en la de cascada, sin embargo, como se mencionó anteriormente, en incremental se puede repetir todo el proceso de forma simultanea varias veces hasta completar los objetivos del proyecto.

1.4.2. Ágiles

Las metodologías de desarrollo de tipo ágil no se tratan de procesos predeterminados, sino hacen énfasis en la colaboración para lograr resultados y cambios evolutivos en los proyectos a los cuales se enfocan. Son más flexibles a los cambios y desarrollos de especificaciones por parte del usuario, por lo que se dedica menos tiempo a la planificación inicial y la priorización de actividades (27).

Algunas de estas metodologías son muy semejantes a la metodología incremental que se describió anteriormente. La diferencia más evidente es que en las metodologías ágiles, las iteraciones son mucho más cortas y rápidas, permitiendo realizar ajustes menores, en lugar de hacer grandes modificaciones al proyecto.

Algunos ejemplos de metodologías ágiles se mencionan a continuación.

1.4.2.1. Kaizen

Kaizen es una metodología de desarrollo que se basa en la idea de que si realizamos pequeñas mejoras de forma continua a lo largo del tiempo, estas pueden conducir a cambios importantes a largo plazo (31). Su objetivo es implementar mejoras y reducir riesgos para construir un entorno eficiente y productivo que agilice el alcance de metas en un proyecto.

Kaizen proviene de la combinación de las palabras japonesas “kai”, que significa cambio, y “zen”, mejorar, por lo que traducción en conjunto se podría definir como “proceso de mejora continua”. Su filosofía se centra en asumir la cultura del mejoramiento continuo, es decir, si se es capaz de hacer algo bien, se puede hacer mejor (32). Además, propone ver los problemas como áreas de oportunidad para mejorar las cosas. Una de sus ventajas es que, no importa cuán grande o pequeño sea un avance, solo busca que las mejoras sean constantes.

Como metodología, proporciona herramientas y recomendaciones que ayudan a

crear un plan de acción para implementar las acciones y mejoras necesarias. Hace más sencillo el proceso de encontrar errores o deficiencias en el proyecto, y también posibilita la opción de solucionar estos contratiempos y llevar un seguimiento estratégico de ellos para evitar que se repitan.

Kaizen está conformado por cuatro etapas que se pueden apreciar en la Figura 1.9.

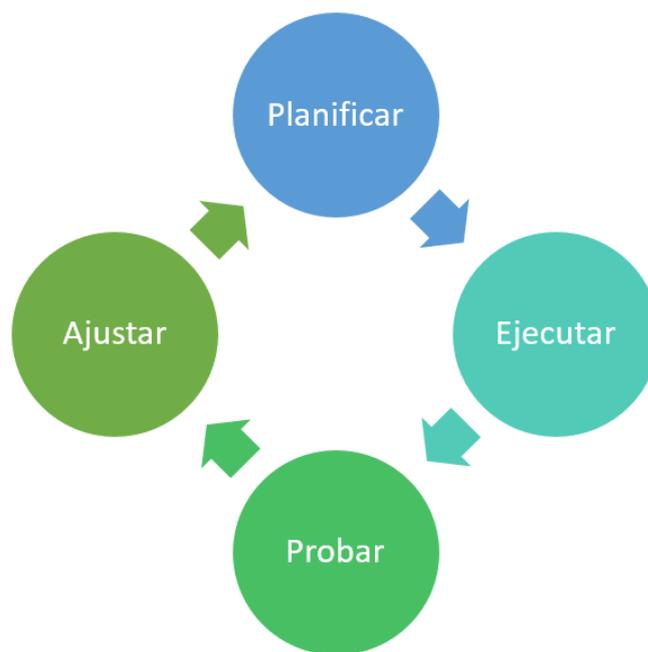


Figura 1.9: Metodología Kaizen (Elaboración propia)

Estas etapas consisten cada una en:

- Planificar: Durante esta primera etapa, se fijan los objetivos, indicadores de éxito, plan y cronograma de actividades, etc.
- Ejecutar: Una vez definidos los detalles de la etapa anterior, en esta se procede a ejecutar el plan de actividades en los tiempos establecidos. Sin embargo, se recomienda hacerlo gradualmente para asegurar un buen funcionamiento en las actividades desarrolladas.
- Probar: Como el nombre lo sugiere, en esta etapa se pone a prueba lo realizado

en la anterior para garantizar que todo vaya según lo planeado en un inicio. Es aquí donde se suelen encontrar los errores cometidos durante el desarrollo y se toma nota de ellos para la siguiente fase.

- Ajustar: Durante esta etapa se realizan los ajustes y las correcciones de los errores detectados en “Probar”. Al tratarse de una metodología cíclica, se puede regresar a “Planeación” de ser necesario para seguir mejorando el proyecto.

1.4.2.2. Kanban

Kanban es un conjunto de mecanismos que son útiles para definir, gestionar y mejorar procesos para llevar a cabo trabajos o actividades en las que interviene la creatividad y el diseño para desarrollar productos de software (33).

Una de sus principales características es el principio de “empieza por donde estés”, el cual permite que la metodología se introduzca en un proyecto ya iniciado de una forma rápida para resolver los problemas de organización acumulados previamente. Esto hace que haya un cambio favorable en el desarrollo del proyecto en camino para cumplir sus objetivos.

Durante la integración de Kanban a un proyecto, los procesos necesarios para lograr los objetivos del mismo se pueden listar en pasos de descubrimiento del conocimiento o del desarrollo de un sistema, visualizados en un tablero Kanban, como el del ejemplo de la Figura 1.10 (33).

El significado literal de Kanban es “tarjeta” o “señal”, por lo que no es de extrañarse que haya un tablero con letreros involucrado en el proceso de desarrollo. Este representa un sistema de flujo en el que los elementos de trabajo fluyen a través de las diversas etapas de un proceso, el orden de izquierda a derecha indica la forma en que se deben de seguir los procesos.

Así mismo, esta es una característica de Kanban que hace posible que trabaje en conjunto con otras metodologías y se complementen mutuamente.



Figura 1.10: Metodología de Kanban (Elaboración propia)

En este trabajo se pretende trabajar con estas dos metodologías ágiles, de tal forma que se complementen para llevar a cabo el objetivo del mismo.

Kanban se puede dividir en tantas etapas como sea conveniente para el proyecto, pero las 3 principales son:

- **Pendiente:** En esta parte se hace la lista de actividades que deben realizarse para cumplir con los objetivos del proyecto.
- **En proceso:** En este apartado se colocan las actividades que se están en proceso en ese momento con el fin de que el trabajo de los desarrolladores sea supervisado y controlado. Es común que los desarrolladores se saturen al hacer varias actividades simultáneamente y esto genera incidencias, por lo que es recomendable poner un límite en el número de actividades que pueden estar en esta etapa a la vez. Además, esto también ayuda a detectar si las actividades están tomando más tiempo del previsto para su desarrollo y tomar las medidas pertinentes.
- **Terminado:** En este apartado se conglomeran las actividades que ya han sido finalizadas. Es muy útil para medir la productividad en el desarrollo del proyecto y vigilar que los tiempos de entrega se estén respetando.

La metodología Kanban es sencilla pero eficaz a la hora de mantener orden durante el proceso de desarrollo. Es por eso que puede ser fácilmente introducida dentro de Kaizen y a su vez, Kaizen puede ser implementado usando principios de Kanban para agilizar sus etapas relacionadas con el cumplimiento de objetivos.

De este modo, ambas metodologías se complementan la una a la otra y brindan una metodología completa, específica y ágil.

Herramientas Utilizadas

2.1. Herramientas de desarrollo

Para la elaboración del sistema propuesto en este trabajo fue necesario hacer uso de diferentes herramientas de desarrollo web y de software. Para poner en contexto al lector, estas herramientas se describen a continuación.

2.1.1. Python

A grandes rasgos, lo que se puede decir de Python es que, es un lenguaje de programación de alto nivel, es decir, es un medio que los humanos pueden usar para comunicarse con las máquinas y así darles instrucciones, pero son más cercanos al entendimiento humano que al de la máquina. Es por eso que está principalmente caracterizado por ser más conciso que otros lenguajes de programación (C, Java, JavaScript, PHP etc.) y requiere de menos esfuerzo por parte del programador para realizar un proceso.

Fue desarrollado por Guido van Rossum y sigue actualizándose constantemente gracias a que cuenta con una gran comunidad de desarrolladores. Lo que lo hace versátil para usarse en cualquier campo de trabajo, desde el análisis de datos hasta la producción de películas, pues al contar con muchas librerías de acceso gratuito, que son conjuntos de archivos con código útil, ahorran muchas horas de trabajo a los desarrolladores.

Otra de sus grandes ventajas es que el software necesario para su interpretación es libre y compatible con diversos sistemas operativos como Microsoft, Apple, Linux, entre otros. Por lo que actualmente se puede implementar en cualquier computadora que tenga la arquitectura y recursos necesarios para soportar su ejecución.

2.1.2. Visual Studio Code

Visual Studio Code (VS Code) es un entorno de desarrollo integrado (IDE) ligero y altamente personalizable desarrollado por Microsoft (34).

Es una herramienta de código abierto y multiplataforma que ofrece soporte para una gran variedad de lenguajes de programación, que incluye a JavaScript, TypeScript, Python, C#, entre otros. VS Code también es conocido por su rendimiento, facilidad de uso, y una gran cantidad de extensiones que permiten a los desarrolladores personalizar su experiencia al programar según sus necesidades específicas.

2.1.3. StarUML

StarUML es una herramienta de modelado de software basado en estándares UML (Lenguaje de Modelado Unificado por sus siglas en inglés) que es un lenguaje gráfico que sirve para generar todo tipo de diagramas en los que se pueda visualizar, documentar y estructurar un sistema (35).

Para este trabajo se hizo uso de esta herramienta para hacer el diagrama de casos de uso en el que se explica de forma general el funcionamiento del sistema.

2.1.4. MockFlow

Mockflow es un software wireframe online en donde se puede diseñar, planear, visualizar y compartir los diagramas para la construcción de un sitio web (36).

Ofrece también a los usuarios una gran biblioteca de componentes mockup, íconos, stickers y otros elementos, para crear una interfaz limpia de una manera rápida y

2. HERRAMIENTAS UTILIZADAS

efectiva. Además, al tratarse de una herramienta en línea, facilita la colaboración en tiempo real con otros desarrolladores, compañeros de equipo o usuario final.

2.1.5. Flask

Flask es una herramienta escrita en Python que ofrece un esquema de trabajo, y una serie de utilidades y funciones que facilitan el desarrollo de aplicaciones web bajo el patrón Modelo-Vista-Controlador (MVC)(37).

Se eligió Flask para el desarrollo del sistema debido a las ventajas que ofrece, sobre todo por la rapidez con la que se puede obtener una aplicación web funcional, además de contar con una flexibilidad que permite integrar funciones propias de Python con otras herramientas necesarias como SQL y HTML

2.1.6. JavaScript

JavaScript es un lenguaje de programación interpretado, lo que quiere decir que no es necesario compilar los programas para ejecutarlos. Dicho de una forma más simple, los programas escritos con JavaScript se pueden ejecutar directamente en cualquier navegador sin necesidad de procesos intermedios (38).

Este lenguaje se introdujo en 1995 como una forma de agregar programas a páginas web en el navegador Netscape Navigator con el propósito de hacer que las páginas web fueran dinámicas (39), es decir que incorporaran efectos visuales llamativos, controlaran las acciones de las páginas web cuando el usuario interaccionara con ellas, etc. El lenguaje ha sido desde entonces adoptado por todos los otros navegadores web y ha hecho que las aplicaciones web modernas sean como se conocen hoy en día.

2.1.7. PostgreSQL

Como mismo PostgreSQL nos dice desde su página oficial (40) es un potente sistema de base de datos relacional de objetos de código abierto con más de 35 años de desarrollo

activo que le ha ganado una sólida reputación por su confiabilidad, solidez de funciones y rendimiento.

Su capacidad para el manejo de los datos es superior en cuanto a la velocidad a la que responde y también que esta no disminuye con el aumento en el tamaño de la base de datos, por lo que su fiabilidad e integridad son muy altas también a pesar de tratarse de una herramienta de código abierto.

2.1.8. Microsoft Azure

Microsoft Azure es una plataforma en la nube que ofrece una gran variedad de servicios para satisfacer las necesidades de infraestructura de empresas y desarrolladores. Dos de sus servicios destacados son el Servidor PostgreSQL y App Service.

El Servidor PostgreSQL de Azure es un servicio de base de datos de tipo relacional que permite a los usuarios implementar, administrar y escalar bases de datos PostgreSQL en la nube de Azure (41). Ofrece características como alta disponibilidad, copias de seguridad automáticas y actualizaciones automáticas para garantizar el rendimiento y la fiabilidad de las aplicaciones.

Además, Azure también cuenta con App Service que es un recurso de plataforma como servicio (PaaS), que facilita a los desarrolladores la creación, implementación y escalado de aplicaciones web y móviles (42). Una de sus ventajas es que es compatible con múltiples lenguajes de programación y marcos. Incluso ofrece funciones como escalado automático, integración continua y administración de identidades para optimizar el desarrollo de aplicaciones.

En conjunto, estos servicios de Azure brindan a los usuarios la flexibilidad, escalabilidad y confiabilidad necesarias para impulsar sus aplicaciones en la nube.

2.2. Algoritmos usados

Dentro de la etapa en la que se hizo el análisis de los datos recolectados, se utilizaron los algoritmos que se describen a continuación.

2.2.1. K-Means

El algoritmo K-Means agrupa los datos tratando de separar muestras en n grupos de igual varianza (medida de dispersión que representa la variabilidad de una serie de datos con respecto a su media), minimizando un criterio conocido como inercia o suma de cuadrados dentro del grupo (43). Este algoritmo requiere que se especifique el número de clústeres. Se adapta bien a un gran número de muestras y se ha utilizado en una amplia gama de áreas de aplicación en muchos campos diferentes.

El algoritmo de K-Means divide un conjunto de N muestras X en K grupos C disjuntos, cada uno descrito por la media μ_j de las muestras en el conglomerado. Los medios se denominan comúnmente como centroides del grupo; Tomando en cuenta que no son, en general, puntos de X , aunque viven en el mismo espacio.

El algoritmo K-means tiene como objetivo elegir centroides que minimicen la inercia, o el criterio de suma de cuadrados dentro del grupo, como se define en la ecuación:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|\mu_j - \mu_i\|^2) \quad (2.1)$$

2.2.2. Método del codo

El método del codo es una técnica utilizada en análisis de clústeres para determinar el número óptimo de clústeres en un conjunto de datos. Consiste en graficar la suma de las distancias cuadradas intra-cluster, también conocidas como inercia, en función del número de clústeres y observar el punto en el que se produce un codo.º un cambio brusco en la pendiente de la curva (44). Este punto sugiere el número óptimo de clústeres,

ya que representa el nivel en el que se añaden más clústeres sin proporcionar mucha ganancia en la reducción de la suma de las distancias cuadradas intra-cluster.

2.2.3. Inercia

La inercia se puede reconocer como una medida de la coherencia interna de los clústeres. Sufre de varios inconvenientes (43):

- La inercia supone que los cúmulos son convexos e isotrópicos, lo que no siempre es así. Responde mal a racimos alargados o variedades con formas irregulares.
- La inercia no es una métrica normalizada: solo se sabe que los valores más bajos son mejores y cero es óptimo. Pero en espacios de dimensiones muy altas, las distancias euclidianas tienden a inflarse (este es un ejemplo de la llamada "maldición de la dimensionalidad"). Ejecutar un algoritmo de reducción de dimensionalidad como el análisis de componentes principales (PCA) antes del agrupamiento de k-medias puede suponer una solución a este problema y acelerar los cálculos.

2.2.4. Análisis de componentes principales (PCA)

PCA es una técnica de reducción de dimensionalidad que permite simplificar la complejidad de los espacios con múltiples dimensiones, mientras conserva su información (45).

Supóngase que existe una muestra con n individuos cada uno con p variables (X_1, X_2, \dots, X_p) , es decir, el espacio muestral tiene p dimensiones. PCA permite encontrar un número de factores subyacentes ($z < p$) que explican aproximadamente lo mismo que las p variables originales. Donde antes se necesitaban p valores para caracterizar a cada individuo, ahora bastan z valores. Cada una de estas z nuevas variables recibe el nombre de componente principal.

PCA se puede usar para sintetizar la información que proporcionan las múltiples variables en unos pocos componentes. Aunque, no se puede pasar por alto la importancia de disponer del valor de las variables originales para calcular esos componentes.

2.2.5. Regresión logística

La regresión logística es un método estadístico utilizado para modelar la relación entre una variable dependiente binaria (dicotómica) y una o más variables independientes (46). A diferencia de la regresión lineal, la regresión logística utiliza una función logística para modelar la probabilidad de que la variable dependiente tome uno de los dos posibles valores. Es ampliamente utilizada en áreas como la medicina, la biología, la ciencia social y el análisis de datos. Existen varias medidas que se pueden usar para analizar los modelos de regresión logística, algunas se mencionan a continuación.

2.2.5.1. Log-Likelihood

En general, en estadística y aprendizaje automático, el Log-Likelihood (log-verosimilitud), es una medida que se usa para evaluar la bondad de ajuste del modelo a los datos observados (47). En otras palabras, se trata de una función que evalúa la probabilidad de observar los datos de entrenamiento dados los parámetros estimados por el modelo. En la regresión logística, se maximiza la log-verosimilitud para encontrar los coeficientes óptimos del modelo.

Es decir que, si el resultado es una medida alta, refiere que el modelo es capaz de explicar mejor los datos observados.

En resumen, la log-verosimilitud en la regresión logística de Python es una medida fundamental para evaluar qué tan bien el modelo se ajusta a los datos observados y se utiliza para ajustar los parámetros del modelo durante el entrenamiento.

2.2.5.2. Pseudo R^2

El pseudo R^2 (también conocido como pseudo R-squared) es una medida de bondad de ajuste que intenta evaluar qué tan bien se ajusta el modelo de regresión logística a los datos observados. A diferencia del R^2 en modelos lineales, los pseudo R^2 en regresión logística no tienen una interpretación directa como la proporción de varianza explicada, pero proporcionan una medida relativa de ajuste en comparación con un modelo nulo (48). Algunas de las métricas pseudo R^2 comunes en regresión logística incluyen el pseudo R^2 de McFadden, el pseudo R^2 de Cox y Snell, y el pseudo R^2 de Nagelkerke.

Estas métricas pseudo R^2 no tienen una interpretación directa como porcentaje de varianza explicada, pero pueden compararse entre modelos para evaluar cuál proporciona un mejor ajuste relativo al modelo nulo.

2.2.5.3. Grados de libertad del modelo (Df_model)

Representan el número de parámetros estimados menos uno (debido a la restricción de que la suma de los residuos sea cero). En el caso de la regresión logística, esto incluirá los coeficientes para las variables predictoras y posiblemente términos adicionales dependiendo de la configuración del modelo (49).

A medida que aumenta df_model, el modelo puede volverse más flexible y capaz de ajustarse mejor a los datos observados, pero también existe el riesgo de sobreajuste si se incorporan demasiadas variables predictoras en relación con el número de observaciones.

2.2.5.4. LL-Null

LL-Null se refiere a la log-verosimilitud del modelo nulo. Este término se utiliza para evaluar la calidad relativa del modelo ajustado en comparación con un modelo de referencia simple que incluye solo la constante y ningún predictor adicional (50).

El LL-Null específicamente mide la log-verosimilitud del modelo que incluye solo al

modelo nulo.

La diferencia entre la log-verosimilitud y el LL-Null proporciona información sobre cuánto mejor (o peor) se adapta el modelo ajustado a los datos en comparación con el modelo nulo. Un valor más alto de LL-Null indica que el modelo ajustado no es significativamente mejor que el modelo nulo.

2.2.5.5. LLR p-value

El LLR (Likelihood Ratio Test) p-value se usa para evaluar la significancia global del modelo en comparación con un modelo nulo. Este valor p se deriva de la comparación de la log-verosimilitud del modelo ajustado con la log-verosimilitud del modelo nulo (46).

El LLR p-value indica la probabilidad de observar una diferencia en las log-verosimilitudes tan grande o más grande que la observada, si el modelo ajustado no es realmente mejor que el modelo nulo. Un LLR p-value pequeño (generalmente menor que un nivel de significancia predefinido, como 0.05) sugiere que el modelo ajustado es significativamente mejor que el modelo nulo.

2.2.5.6. Coeficientes estimados (Coef)

Se trata de los coeficientes estimados de cada una de las variables independientes en el modelo. Estos coeficientes indican cómo cada variable independiente afecta el logaritmo de la odds ratio (log-odds) de la variable dependiente binaria (51).

Tener por resultado un coeficiente positivo refiere que un aumento en el valor de la variable independiente está asociado con mayores probabilidades de que la variable dependiente sea 1 (en lugar de 0), manteniendo constantes las demás variables en el modelo, mientras que un coeficiente negativo indica que un aumento en el valor de la variable independiente está asociado con menores probabilidades de que la variable dependiente sea 1 (en lugar de 0), manteniendo constantes las demás variables.

2.2.5.7. Error estándar (std err)

El error estándar (std err) de los coeficientes estimados se usa para medir la precisión con la que se han estimado esos coeficientes. Da a conocer cuánto se espera que varíen los coeficientes si el modelo se ajusta repetidamente a diferentes muestras de datos del mismo tamaño de la población (52).

Al analizar esta medida, se puede decir que un error estándar pequeño sugiere una mayor precisión en la estimación del coeficiente correspondiente. Por otro lado, un error estándar grande indica una mayor incertidumbre en la estimación del coeficiente.

2.2.5.8. Valor de Z

El valor z se refiere a la estadística z asociada a los coeficientes estimados de las variables independientes en el modelo (53). Los coeficientes de regresión estimados se asocian con una estadística z y un valor p que indican la significancia de cada coeficiente en relación con la variable que se está prediciendo.

En términos generales, el valor z para un coeficiente específico se calcula dividiendo el coeficiente estimado por su error estándar. Este valor z sigue una distribución normal estándar bajo la hipótesis nula de que el coeficiente es igual a cero, es decir, la variable independiente no tiene efecto sobre la variable dependiente. Por el contrario, si z tuviera un valor absoluto alto, indicaría que la variable independiente es significativamente positiva o negativa (según sea el caso) para la variable dependiente.

2.2.5.9. $P > |z|$

En Python, al trabajar con modelos de regresión logística, la expresión $P > |z|$ se refiere al valor p asociado a la estadística z de los coeficientes estimados en el modelo (54). Este valor p indica la probabilidad de observar el efecto estimado del coeficiente bajo la hipótesis nula de que no hay efecto, o, dicho de otra forma, que el coeficiente

es cero.

Si $P > |z|$ fuera bajo, se rechaza la hipótesis nula, lo que sugeriría que el coeficiente es estadísticamente significativo y viceversa, si $P > |z|$ fuera alto, no se podría rechazar la hipótesis nula, lo que indicaría que el coeficiente no es estadísticamente significativo.

2.2.5.10. Límites del intervalo de confianza

Generalmente se refiere a los límites del intervalo de confianza al 95 % para los coeficientes estimados de un modelo. Estos valores indican el rango dentro del cual se espera que esté el valor real del coeficiente con un nivel de confianza del 95 % (55).

Los números [0.025 0.975] representan los límites inferior y superior del intervalo de confianza del 95 % para un coeficiente específico en el modelo de regresión logística.

El valor 0.025 se refiere al límite inferior del intervalo de confianza del 95 %. Esto significa que con un 95 % de confianza, se espera que el verdadero valor del coeficiente sea mayor o igual a este valor. Por su lado, el valor 0.975 indica el límite superior del intervalo de confianza del 95 %. Con un 95 % de confianza, se espera que el verdadero valor del coeficiente sea menor o igual a este valor.

Si el intervalo [0.025 0.975] no incluye el valor cero, entonces el coeficiente correspondiente se considera estadísticamente significativo. Esto implica que hay evidencia suficiente para determinar que el efecto de la variable independiente es diferente de cero en la variable dependiente.

Los números [0.025 0.975] representan los límites inferior y superior del intervalo de confianza del 95 % para un coeficiente específico en un modelo predictivo.

2.2.5.11. Likelihood Ratio Chi-square

La prueba Likelihood Ratio Chi-square (LR Chi2) en regresión logística es una medida estadística utilizada para evaluar la significancia global del modelo. Esta prueba compara la verosimilitud del modelo ajustado con la verosimilitud de un modelo más

simple, generalmente uno que no incluye ninguna variable predictora (modelo nulo) (46).

La forma de interpretar esta medida del modelo, es observar si el valor de LR Chi2 es elevado. De ser así, aunado a un valor p asociado pequeño indica que el modelo completo se ajusta significativamente mejor que el modelo nulo, lo que sugiere que al menos una de las variables predictoras tiene un efecto significativo en la variable dependiente.

2.2.6. Regresión lineal

La regresión lineal es un método estadístico utilizado para modelar la relación entre una variable dependiente y una o más variables independientes mediante una función lineal (56). El objetivo es encontrar la mejor línea recta que se ajuste a los datos observados. Se utiliza en diversas áreas, desde análisis de datos hasta predicciones.

Algunas de las medidas que se utilizan para evaluar un modelo de regresión lineal son las que se explican a continuación.

2.2.6.1. Coeficiente de determinación R^2

El coeficiente de determinación R^2 , o R-squared, es una medida estadística que denota qué tan bien se ajustan los datos a la línea de regresión ajustada en un modelo de regresión lineal (56).

El valor de R^2 varía entre 0 y 1, por lo que un valor más cercano a 1 indica que una mayor proporción de la varianza en la variable dependiente está explicada por el modelo de regresión. Si $R^2 = 1$ significa que el modelo ajusta perfectamente los datos. Por el contrario, si $R^2 = 0$ significa que el modelo no demuestra ninguna variabilidad en los datos.

2.2.6.2. R^2 ajustado (Adj. R-squared)

El R^2 ajustado es una versión mejorada del coeficiente de determinación R^2 . A diferencia de R^2 , que aumenta o no cambia cada vez que se añade una variable predictora al modelo, el R^2 ajustado penaliza la inclusión de variables irrelevantes que no mejoran significativamente la capacidad predictiva del modelo (56).

R^2 ajustado es una medida que penaliza modelos que tienen muchas variables predictoras, especialmente aquellas que no mejoran significativamente la capacidad predictiva del modelo.

Comparado con R^2 , R^2 ajustado es típicamente más bajo, especialmente cuando se añaden variables que no contribuyen suficientemente a la explicación de la varianza en la variable dependiente, por el contrario, un valor alto de R^2 ajustado indica que una mayor proporción de la varianza en la variable dependiente está explicada por el modelo ajustado, teniendo en cuenta el número de variables predictoras incluidas.

2.2.6.3. Estadístico F (F-statistic)

El estadístico F (también conocido como F-statistic) se utiliza para evaluar la significancia global del modelo de regresión (49).

Si el valor p, que es la probabilidad de observar un estadístico F tan extremo como el observado si la hipótesis nula fuera cierta, es menor que un nivel de significancia predeterminado, se rechaza la hipótesis nula (que afirma que no hay relación entre las variables predictoras y la variable dependiente) en favor del modelo completo. Por lo tanto, al menos una de las variables predictoras en el modelo completo tiene una relación significativa con la variable dependiente.

2.2.6.4. Prob (F-statistic)

Esta medida se refiere al valor p asociado con la estadística F calculada para probar la significancia global del modelo de regresión.

En la regresión lineal, la estadística F se utiliza para determinar si al menos una de las variables independientes tiene un efecto significativo sobre la variable dependiente. Dicho de otra forma, prueba la hipótesis nula de que todos los coeficientes de regresión son cero, frente a la hipótesis alternativa de que al menos uno de ellos es distinto de cero (57).

Por tanto, el valor p asociado con la estadística F calculada señala la probabilidad de obtener un valor de la estadística F al menos tan extremo como el observado, bajo la hipótesis nula de que no hay efecto significativo de las variables explicativas en la variable dependiente. Un valor p bajo sugiere que al menos una de las variables explicativas tiene un efecto significativo en la variable dependiente(58).

Es decir, si Prob (F-statistic) es menor que un nivel de significancia dado, se rechaza la hipótesis nula en favor del modelo completo. Esto sugiere que al menos una de las variables predictoras contribuye significativamente a explicar la variabilidad en la variable dependiente y si Prob (F-statistic) es mayor que el nivel de significancia, no se rechaza la hipótesis nula, lo que indica que el modelo completo no proporciona una mejora significativa sobre el modelo nulo.

2.2.6.5. Akaike Information Criterion (AIC)

El AIC es un criterio de información utilizado para la selección de modelos en el contexto de regresión lineal y otros modelos estadísticos. Esta medida penaliza el sobreajuste al modelo al tomar en cuenta tanto la bondad de ajuste del modelo como la complejidad del mismo (59).

El AIC se utiliza para comparar modelos alternativos. Entre dos modelos, se prefiere

aquel que tenga el valor de AIC más bajo, ya que apunta un mejor equilibrio entre ajuste y complejidad, ya que, demuestra una mejor capacidad predictiva y generalización en datos futuros.

2.2.6.6. Bayesian Information Criterion (BIC)

BIC, también conocido como Criterio de Información de Schwarz, es una métrica utilizada para la selección de modelos. Es útil para comparar diferentes modelos ajustados a un mismo conjunto de datos y elegir el que mejor balancee la calidad del ajuste con la complejidad del modelo (60).

La forma correcta de interpretar la métrica BIC es en comparación con otros modelos, no en términos absolutos, ya que, se debe tomar en cuenta la competencia entre modelos para evitar interpretaciones erróneas y decisiones subóptimas.

Un modelo con un BIC más bajo se considera mejor que otros con valores más altos, ya que sugiere un buen ajuste con una penalización menor por la complejidad del modelo.

2.2.6.7. Estadístico t

Es un estadístico utilizado para evaluar la significancia de cada coeficiente del modelo.

Su función es probar la hipótesis nula de que el coeficiente de una variable independiente es igual a cero (61) (es decir, que la variable no tiene efecto en la variable dependiente).

Un valor t alto (en valor absoluto) generalmente indica que el coeficiente es significativamente diferente de cero, lo cual indica que la variable tiene gran influencia sobre la variable dependiente.

2.2.6.8. $P > |t|$

El valor $P > |t|$, también conocido como valor p asociado al estadístico t, es una medida que indica la probabilidad de observar un valor del estadístico t tan extremo o más extremo que el valor calculado, bajo la hipótesis nula de que el coeficiente es igual a cero (62). En otras palabras, el valor p mide la evidencia en contra de la hipótesis nula.

Un valor $P > |t|$ bajo apunta a que hay suficiente evidencia para rechazar la hipótesis nula. Esto significa que el coeficiente es significativamente diferente de cero y, por ende, la variable independiente tiene un efecto significativo en la variable dependiente.

Por el contrario, un valor alto supone que no hay suficiente evidencia para rechazar la hipótesis nula. Esto significa que el coeficiente no es significativamente diferente de cero, y la variable independiente puede no tener un efecto importante en la variable dependiente.

2.3. Herramientas de análisis estadístico de datos

El análisis estadístico es una parte fundamental en una investigación, pues aporta información que no es evidente a primera vista y ayuda a la resolución de problemas relacionados con los datos recolectados.

Las medidas estadísticas que van a ser utilizadas en la realización del sistema propuesto se presenten a continuación.

2.3.1. Máximo y mínimo

Como el correspondiente nombre lo sugiere, el valor máximo, es el valor más grande dentro de un conjunto de datos. Por el contrario, el valor mínimo, es el más pequeño (63).

2. HERRAMIENTAS UTILIZADAS

Ambos datos se pueden obtener al comparar cada dato del conjunto, buscando el más grande y el más pequeño respectivamente.

Conocer estos valores sirven en el análisis de los datos para conocer de qué manera están distribuidos los datos de la muestra.

2.3.2. Media aritmética

En (64), se define a la media aritmética o promedio (\bar{x}) como el resultado de sumar n veces un conjunto de n mediciones y dividirlo entre n .

Se puede apreciar de una mejor manera en la ecuación:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.2)$$

donde x es cada medición del conjunto.

Generalmente, esta medida se utiliza cuando es necesario "resumir" un conjunto de datos numéricos. Pues representa el equilibrio entre los valores del conjunto. En otras palabras, es el valor que tendrían las mediciones, si todas ellas fueran iguales o el que le correspondería a cada una de las mediciones del conjunto si su suma total se repartiera por igual.

2.3.3. Desviación estándar

En (65), se explica que la desviación estándar (S) es una medida de extensión o variabilidad en la estadística descriptiva. Se utiliza para calcular la variación o dispersión en la que los puntos de datos individuales difieren de la media.

La ecuación que expresa la desviación estándar es la que se muestra a continuación:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (2.3)$$

donde x es el valor de cada medida, \bar{x} es la media aritmética y n es la cantidad de

valores de la muestra.

2.3.4. Rango intercuartílico

Un cuartil (Q) es cada uno de los tres valores que pueden dividir un grupo de números ordenados de menor a mayor en cuatro partes iguales (66).

Entonces cuando se habla de un rango intercuartilico (IQR) se hace referencia a la medida de dispersión de un conjunto de datos que expresa la diferencia o la distancia entre el cuartil superior e inferior.

Esto se expresa en la siguiente ecuación:

$$IQR = Q_3 - Q_1 \quad (2.4)$$

donde Q_3 es el cuartil superior y Q_1 es el cuartil inferior.

2.3.5. Prueba de bondad de ajuste de Kolmogórov-Smirnov

La prueba de bondad de ajuste de Kolmogorov-Smirnov, es un procedimiento de que se le aplica a una muestra y permite medir el grado de concordancia existente entre la distribución de un conjunto de datos y una distribución teórica específica (67). El propósito de esta prueba es averiguar si los datos provienen de una población que tiene la distribución teórica especificada, es decir, contrasta si las observaciones podrían razonablemente proceder de la distribución especificada.

Dentro de los análisis estadísticos varias de las pruebas paramétricas requieren que las variables se distribuyan de forma normal. Esta prueba puede usarse con el fin de saber si una variable de una muestra se distribuye de forma normal.

2.3.6. Correlación lineal

La correlación lineal es un método estadístico que permite cuantificar la relación lineal existente entre dos variables. Existen varias funciones estadísticas llamadas coeficientes de correlación lineal, creados con el propósito de medir este tipo de asociación, algunos de los más usados son Pearson y Spearman (68).

Pearson: Este coeficiente de correlación funciona bien con variables cuantitativas que tienen una distribución normal o próxima a la normal. Es más sensible a los valores extremos que las otras dos alternativas.

Spearman: Se emplea con variables cuantitativas. En lugar de utilizar directamente el valor de cada variable, los datos son ordenados y reemplazados por su respectivo orden ranking. Es un método no paramétrico muy utilizado cuando no se satisface la condición de normalidad necesaria para aplicar la correlación de Pearson.

Estado del Arte

En este capítulo se presentan trabajos encontrados en la literatura relacionados a la problemática abordada en esta investigación. Posteriormente, se analizan estos trabajos y se realiza una comparación con el propuesto.

3.1. Un sistema informatizado de apoyo a las decisiones clínicas para el manejo de la depresión en atención primaria

Aragones et al. describen en este trabajo los fundamentos y principales características de una guía clínica informatizada para el manejo de la depresión mayor, desarrollada en el sistema sanitario público de Cataluña (69).

El resultado descrito en este trabajo es un sistema que pretende ser un apoyo para establecer diagnósticos de depresión fiables y precisos, con el fin de elegir el tratamiento idóneo a partir de las características de la enfermedad y del propio paciente. Además, enfatizan en la importancia de la detección del riesgo de suicidio y de un plan de seguimiento sistemático, para evaluar la evolución clínica del paciente y de ese modo, poder adecuar las intervenciones terapéuticas a las necesidades de cada momento.

Esta herramienta propone el uso de algoritmos inteligentes, los autores no especi-

can cuales, pero se puede deducir que fueron del tipo supervisado, que contemplan los múltiples escenarios posibles del estado en el que se encuentra un paciente. Apoya al médico ofreciendo recomendaciones dinámicas en función del contexto clínico de cada paciente y se utiliza de forma síncrona durante la visita del paciente, permitiendo llevar la atención en tres módulos: diagnóstico, tratamiento y seguimiento, con el propósito de brindar un servicio oportuno y completo.

3.2. How Facebook AI Helps Suicide Prevention

A nivel mundial, ocurre una muerte por suicidio cada 40 segundos y es la segunda causa principal de muerte entre personas de 15 a 29 años.

Al tratarse de una de las redes sociales más usadas, Facebook es una plataforma donde se comparte una inmensurable cantidad de contenido de todo tipo, por lo que no es de extrañar que cuando alguien está expresando pensamientos de suicidio, haga publicaciones al respecto y es importante conseguirles ayuda lo antes posible.

Ya que los amigos y la familia están conectados a través de Facebook, esta plataforma puede ayudar a una persona en apuros a ponerse en contacto con personas que puedan brindarle apoyo. Desde hace varios años, quienes ven publicaciones que indican que alguien está pensando en suicidarse, tienen la opción de denunciarlas, de ese modo un equipo especializado revisaría el contenido publicado y tomaría las medidas necesarias para salvaguardar la vida de la persona en riesgo. Lamentablemente, la cantidad de datos que deben ser revisados, rebasan la capacidad humana para darles la atención que requieren.

Con la tendencia del uso de modelos de Aprendizaje Automático para agilizar procesos que requieran de un análisis de grandes cúmulos de datos, Facebook implementó un algoritmo de aprendizaje random forest (70). Se trata de un modelo Aprendizaje Automático que se especializa en datos numéricos. Utiliza diversos árboles de decisión

y genera la predicción media de los árboles individuales para analizar texto y aprende tanto de las denuncias hechas por los usuarios como de las decisiones tomadas por el equipo especializado de Facebook, para ayudar a la detección de publicaciones que delatan intenciones suicidas.

Además de llegar al mismo resultado de la OMS acerca de que el sector de la población que más cometía suicidio, eran hombre entre los 15 y 29 años de edad, descubrieron que las primeras horas de la mañana y los domingos, cuando se acerca la semana laboral/escolar, pueden ser momentos comunes para contemplar el suicidio.

El resultado más sobresaliente del equipo de desarrollo fue lograr que el clasificador de Aprendizaje Automático lograra diferenciar entre las publicaciones que verdaderamente denotaban intenciones suicidas y las que no. Esto fue gracias a que se dieron cuenta de que tenían un conjunto más pequeño y preciso de ejemplos negativos, i.e. el conjunto de publicaciones de Facebook que las personas habían marcado potencialmente como conteniendo que denotaba pensamientos suicidas, pero que los revisores capacitados de Operaciones Comunitarias determinaron que no demostraban comportamientos reales donde una persona estuviera en riesgo de cometer autolesiones. La identificación de este conjunto, condujo a un mejor entrenamiento de los clasificadores para identificar auténticas expresiones suicidas.

3.3. Suicide and Suicide Attempt Descriptors by Multimethod Approach

En trabajo de Bojan et al (71), usaron la información recopilada por la Clínica Psiquiátrica Universitaria de Ljubljana en el Registro de Suicidios en la República de Eslovenia, como punto de partida. En la parte computacional, aplicaron una combinación de variables, cuya variación radica dentro de una variable y entre variables. En otras palabras, el comportamiento de una, repercute en el de las demás.

3. ESTADO DEL ARTE

De acuerdo al autor, debido a la limitación del uso de métodos y análisis estadísticos estándar en tales casos, aplicaron una estrategia conocida como: Aproximación Híbrida Multimétodo, que permitió combinar árboles de decisión, algoritmos genéticos y vectores suplementarios, para superar las desventajas y limitaciones que cada algoritmo tiene por sí solo. La investigación incluyó a 56712 personas que intentaron suicidarse y 21913 personas que se suicidaron. En la variable de salida había solo dos tipos de resultado: intento de suicidio y suicidio.

Con base en el análisis del Aprendizaje Automático, definieron atributos de la acción suicida en cuanto a su efecto letal: intento de suicidio y compromiso suicida. El registro de suicidios llevado durante los últimos 40 años muestra el ahorcamiento como el método suicida más utilizado. Los hombres lo emplean más con el propósito de provocar la muerte por autolesión que por un intento de suicidio. Por otro lado, para el sexo femenino el uso de medicamentos es más común para intentar acabar con su vida voluntariamente.

Ante el hecho de que el tipo de lesión auto infringida no ayuda a predecir la muerte del individuo por esa causa, examinaron diferentes formas de cometer suicidio para predecir con mayor precisión el vínculo entre la forma y el resultado de la acción. El modelo de aprendizaje automático aplicado en el trabajo confirmó que los atributos de las formas de autolesión se relacionan con sus diferentes resultados. El modelo en cuestión, es útil en el procesamiento de grandes bases de datos ya que permite que el peso de una variable encuentre la relación en términos de resultado y significado.

3.4. Recognizing states of psychological vulnerability to suicidal behavior: a Bayesian network of artificial intelligence applied to a clinical sample

El estudio realizado por Barros et al (72), se centró en identificar relaciones de dependencia condicional de variables significativas para detectar personas con vulnerabilidad psicológica que indicara si eran propensas a cometer autolesiones con propósitos fatales. Desarrollaron y aplicaron una red bayesiana para establecer esas relaciones de dependencia condicional entre variables para cada uno de los casos estudiados. Estas dependencias condicionales representaron los diferentes estados en los que pueden estar los pacientes en relación con la conducta suicida. El tamaño de la muestra usada fue de 650 pacientes que presentaron síntomas que indicaban trastornos en el estado de ánimo y ansiedad. Los resultados señalaron que las variables dentro de la red bayesiana son parte del estado de vulnerabilidad psicológica de cada paciente y pueden influir en tales estados. Además, descubrieron que esas variables coexisten y son relativamente estables en el tiempo. Los resultados les permitieron crear una herramienta que fuera capaz de detectar estados de vulnerabilidad psicológica asociados al riesgo de suicidio. Evaluaron a los pacientes haciendo uso de la mencionada herramienta que no solo fue capaz de mostrar los factores de riesgo, sino también los protectores. Esto con el fin de orientar a los evaluadores y profesionales encargados de atender a esas personas a señalar aspectos de importantes para la intervención psicoterapéutica. Clasificaron los perfiles de los pacientes en relación a los siguientes aspectos:

- Sentimientos de satisfacción/insatisfacción con la vida.
- Estado de satisfacción/insatisfacción con uno mismo y los logros.
- Razones para vivir/para seguir con vida si uno está pensando en intentar suicidarse.

Los autores resaltaron la necesidad de, en el futuro, hacer más pruebas con la herramienta que desarrollaron en escenarios diferentes para tener información más rica que ayude a los profesionales a saber cuándo intervenir para prevenir conductas suicidas.

3.5. Suicide risk configuration system in a clustered clinical sample: a generalized linear model obtained through the LASSO technique

El objetivo del trabajo realizado por Maino et al. (73) fue identificar factores clínicos y sociodemográficos que aumentarían o disminuirían el riesgo suicida en una muestra clínica de pacientes que buscaron atención médica en el área de salud mental.

Realizaron un estudio transversal en tres centros de Salud de Santiago de Chile. Obtuvieron los datos al hacerle diferentes entrevistas relacionadas con la salud mental del paciente y su situación sociodemográfica. La muestra fue de 544 pacientes, de los cuales 333 presentaban conducta suicida, mientras que los 211 restantes, no.

Después de que hicieron un análisis de agrupamiento jerárquico, los sujetos de la muestra fueron separados en distintos grupos de acuerdo a sus similitudes en cuanto al riesgo que tenían de cometer suicidio. Seguido de este análisis, realizaron otro de tipo regresivo, haciendo uso de una técnica conocida como Operador de Selección y Contracción Mínima Absoluta (LASSO, por sus siglas en inglés) e identificaron factores que disminuyen o aumentan el riesgo de cometer suicidio para cada grupo.

Los resultados de este artículo, revelaron que los grupos estaban separados principalmente por la edad de los pacientes. Así también como que el factor protector más influyente fue tener confianza en la propia resiliencia. Del mismo modo, los factores de riesgo que deben ser vigilados con mayor cuidado, fueron el trastorno depresivo mayor, poco control sobre el manejo de la ira y problemas en las relaciones interpersonales.

Por último, observaron experimentalmente que las actitudes suicidas y el riesgo

de cometer autolesiones, se manifiestan de forma diferente dependiendo de la edad de quien las experimenta y por lo tanto existen diferentes tipos de factores que pueden proteger o aumentar el riesgo de suicidio.

3.6. Identification of Suicide Attempt Risk Factors in a National US Survey Using Machine Learning

La finalidad de García de la Garza et al (74) al hacer este trabajo, fue identificar los factores de riesgo que delatarán futuros intentos de suicidio en la población general utilizando un enfoque de Aprendizaje Automático. El estudio está basado en datos que incluyen más de 2500 preguntas de una extensa encuesta a nivel nacional de adultos estadounidenses.

Los datos se obtuvieron específicamente de la Encuesta Epidemiológica Nacional sobre Alcohol y Condiciones Relacionadas (NESARC, por sus siglas en inglés), realizada en los periodos 2001 a 2002 y 2004 a 2005, lo que dio un resultado de 34,653 entrevistas.

Entrenaron un random forest equilibrado mediante validación cruzada para desarrollar un modelo de riesgo de intento de suicidio. Usaron la predicción del modelo fuera de pliegue para evaluar el rendimiento del modelo, incluida el área bajo la curva del operador del receptor, la sensibilidad y la especificidad.

El modelo arrojó como resultado que los factores más influyentes para determinar si una persona está en riesgo de intentar terminar con su vida, eran el antecedente de ideación o comportamiento suicida, estado de ánimo decaído, realización de actividades con menor cuidado o bajo rendimiento a causa de problemas emocionales, edad joven, bajo desempeño escolar y la crisis financiera de la época.

Finalmente, los autores estimaron que los resultados de su investigación podrían ser de ayuda para el área médica que atiende la salud mental, al servir de guía en la evaluación clínica de los pacientes y dar un seguimiento al desarrollo de niveles más

elevados de riesgo de suicidio.

3.7. A predictive model for automatic detection of loneliness and social isolation using machine learning

El trabajo realizado por (75) tiene por objetivo desarrollar un modelo predictivo de soledad y aislamiento social a partir de la información obtenida del seguimiento de las actividades diarias de un adulto mayor.

Los datos que se utilizaron para la realización del estudio fueron obtenidos del Consejo Nacional de Población (CONAPO) en el 2016 y de otra investigación realizada por Sánchez-Gómez (76).

Hicieron uso de los algoritmos AdaBoost con Random Forest para el modelo de soledad y el algoritmo k-Nearest-Neighbor para el de aislamiento social.

Como resultado obtuvieron un modelo capaz de predecir la soledad y el aislamiento social en adultos mayores basado en unir modelos específicos de soledad y aislamiento social, y en detectar los atributos sobresalientes que ambos modelos de predicción tienen en común.

3.8. Physical frailty prediction model for the oldest old

El propósito del estudio realizado por (77) fue desarrollar modelos predictivos de condiciones de fragilidad en adultos mayores utilizando diferentes métodos de Aprendizaje Automático basados en una base de datos de características clínicas y factores socioeconómicos.

Los datos utilizados fueron obtenidos de una base de datos administrativa de salud que contiene 1.095.612 registros de adultos mayores de 65 años o más con 58 variables de entrada y 6 variables de salida.

Al hacer el estudio de estos datos, usaron una red neuronal artificial, programación genética, máquinas de vectores de soporte, bosque aleatorio, regresión logística y árboles de decisión.

Finalmente obtuvieron modelos de Aprendizaje Automático capaces de predecir condiciones de fragilidad en adultos mayores

3.9. Discusión

De acuerdo a los trabajos revisados y como se puede apreciar en la Tabla 3.1, en todos los trabajos presentados en esta investigación del estado del arte, se hace uso de algoritmos de Aprendizaje Automático. Sin embargo, no todos tuvieron como fin de su investigación contribuir con la detección o diagnóstico de que un paciente perteneciente a la población geriátrica tenga intenciones suicidas y/o de hacer un análisis de los factores de riesgo y/o protectores, que son aquellos que indican cuáles son los motivos o eventos que contribuyen a que una persona sea vulnerable a tomar la decisión de poner fin a su propia vida o lo protejan de tomar tal decisión. Así también, solo dos de ellos tuvieron la iniciativa de desarrollar un sistema para el uso de los especialistas en el área de la salud para facilitar su trabajo. La propuesta que se presenta en este documento, pretende atender todos estos aspectos y hacer un trabajo de investigación y desarrollo más completo que los vistos durante este capítulo, además de apoyar a los especialistas a la recolección de datos.

3. ESTADO DEL ARTE

Artículos	Implementación de algoritmos de AA	Ayuda a la pre-dicción o diagnóstico del suicidio	Genera un sistema de apoyo a especialistas	Análisis de factores			Acceso online al sistema	Se pueden realizar los análisis dentro del sistema	Ayuda al especialista a adquirir los datos	Dirigido a población geriátrica
				Psicológicos	Físicos	Sociales				
(69)	si	si	si	si	si	si	no	si	si	no
(70)	si	si	no	si	no	si	no	no	no	no
(71)	si	no	no	si	si	si	no	no	no	no
(72)	si	no	no	si	si	si	no	no	no	no
(73)	si	si	si	si	si	si	no	si	no	no
(74)	si	si	no	si	si	si	no	no	no	no
(75)	si	si	no	si	si	si	no	no	no	si
(77)	si	no	no	no	si	no	no	no	no	si
Propuesto	si	si	si	si	si	si	si	si	si	si

Tabla 3.1: Tabla comparativa de trabajos relacionados

Metodología de la Investigación

Se propone fusionar las metodologías Kanban y Kaizen con la finalidad de obtener una que permita el desarrollo del proyecto. La metodología resultante es de tipo ágil y consiste en llevar a cabo tres fases principales que, a su vez, contienen sus propias subfases, así como se puede apreciar en la Figura 4.1. A continuación, se detalla en qué consiste cada fase de la metodología planteada.

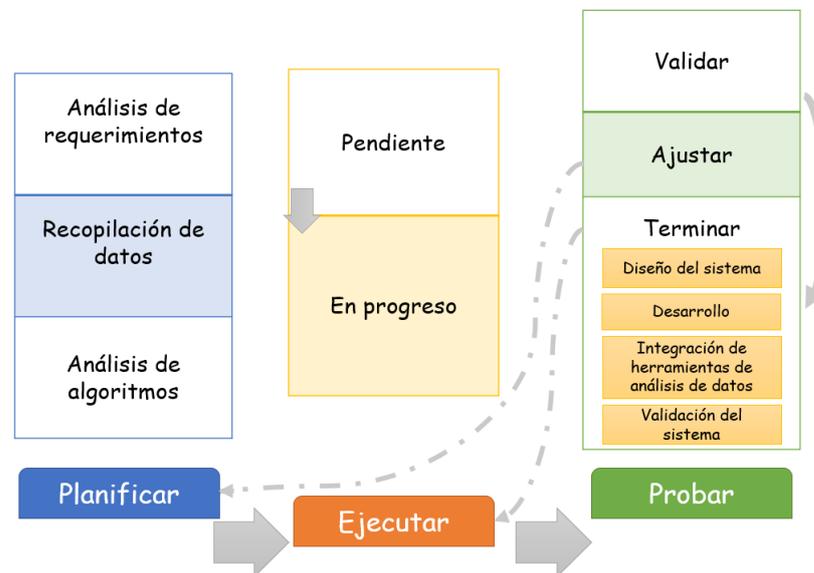


Figura 4.1: Metodología Kaizen-Kanban propuesta (Elaboración propia)

4.1. Planificar

Durante esta fase se hacen todos los preparativos necesarios para empezar a desarrollar la herramienta computacional, que se pueden dividir en tres subfases:

- **Análisis de requerimientos:** La subfase consiste en hacer un estudio detallado de las necesidades que debe satisfacer la herramienta computacional y enlistar las actividades que se deben desarrollar para cumplirlas.
- **Recopilación de datos:** Se hace la integración, selección, limpieza y transformación de los datos obtenidos para realizar las actividades que dependen de ellos para realizar las actividades relacionadas para la solución del problema.
- **Análisis de algoritmos:** Es necesario hacer un análisis de los diferentes algoritmos de Aprendizaje Automático que existen para poder elegir el que mejor se adapta al problema del estudio de los factores de riesgo en el suicidio en México.

4.2. Ejecutar

Dando continuidad a la etapa anterior, en esta se hace el desarrollo de las actividades marcadas para el cumplimiento de los requerimientos de la herramienta computacional. Con el propósito de llevar un control en la realización de estas actividades y haciendo caso a la metodología Kanban, esta fase se divide en:

- **Pendiente:** En esta parte se conglomeran todas las actividades que aún no se han realizado y están en espera. Durante el desarrollo, ayuda a no pasar por alto ninguna actividad y está fuertemente ligada con otra subfase llamada “Terminado” (Que se explica más adelante en Probar) para la detección oportuna de incidencias.

- En progreso: Se colocan las actividades que se están realizando actualmente y permite la visualización del flujo de trabajo. Dichas actividades son aquellas que tratan sobre el desarrollo y se plantearon en la planeación, como el diseño del sistema y la integración de la inteligencia computacional. Kanban recomienda que se fije un límite de actividades que pueden estar simultáneamente en esta subfase para evitar saturación en el desarrollo y por tanto la generación de inferencias propias de esa causa. Una vez completadas las actividades, pasan a la fase siguiente: "Probar".

4.3. Probar

En esta fase se hacen las pruebas de las funciones de la herramienta y se determina si cumple el objetivo planteado. Las subfases a ejecutar en esta fase son:

- Validar: Durante esta subfase se hacen las pruebas al software de dos maneras, la primera tiene que ver con el funcionamiento individual de cada parte del software y la segunda con su ejecución en conjunto con las demás. De ese modo, con estas pruebas, se determina si tales partes son validadas o no. Dependiendo de ese resultado, se pasa a alguna de las siguientes etapas. A "Ajustar" si tuvieron alguna complicación o a "Terminar" si funcionaron correctamente.
- Ajustar: A esta fase pasan las funciones de la herramienta que no se están ejecutando correctamente o que pueden mejorarse. Si es posible hacer esos arreglos, se pasan a la siguiente subfase descrita en este apartado. Si no es así, se puede volver a enviar a la fase "Planificar" para hacer una reestructuración de dicha función para adaptarla a las posibilidades del desarrollo.
- Terminar: Finalmente, a esta subfase pasan las funciones que ya funcionan correctamente y no necesitan más ajustes. Es conveniente vigilar esta subfase junto

4. METODOLOGÍA DE LA INVESTIGACIÓN

con "Pendiente" para supervisar que el desarrollo se esté llevando de la manera planificada y se estén cumpliendo con las fechas de entrega establecidas en el cronograma. Si no fuera así, detectar a tiempo las anomalías que ocasionan esos retrasos y corregirlas oportunamente.

Planeación de la Construcción de la Herramienta

Durante esta primera etapa del desarrollo del sistema, como se habló en el capítulo anterior, se realizaron diferentes actividades para conocer las labores que se espera se realicen con apoyo de dicho sistema, así como los problemas que pueden surgir y anticiparse a ellos.

En las siguientes secciones se detallan las actividades previas a comenzar el desarrollo.

5.1. Análisis de requerimientos

Las especialistas con las cuales se trabajó en conjunto para este proyecto, son la Dra. en Psicología Lilian Elizabeth Bosques Brugada y la Maestra Christian Jazmín Flores Pérez (Licenciada en Gerontología y Maestra Ciencias Biomédicas y de la Salud).

Durante esta actividad se tuvo una discusión sobre el trabajo de las especialistas, como solían hacerlo y los problemas que han tenido hasta el momento. De ese modo hubo la oportunidad de brindarles algunas opciones para solucionar algunas de esas adversidades usando una herramineta computacional y al final se concluyó que las características más indispensables que debe poseer esta herramienta son:

5. PLANEACIÓN DE LA CONSTRUCCIÓN DE LA HERRAMIENTA

- Portabilidad entre diferentes plataformas. Es indispensable el uso en equipo de cómputo y dispositivos móviles.
- Actualización de los datos en tiempo real.
- Acceso de múltiples usuarios a la vez desde diferentes lugares.
- Adquisición de datos de forma independiente, de acuerdo al tipo de información del adulto mayor, para agilizar el proceso.
- Interfaces responsivas de los formularios con la finalidad de que sean amigables con la vista de un adulto mayor promedio.
- Apartado para firmar digitalmente un documento de “Conocimiento informado” (A).
- Disponibilidad de poder ver la información de un adulto mayor en ciertos apartados.
- Registro adecuado de los valores de las respuestas a las preguntas directas e inversas.

5.2. Diseño de la base de datos

La base de datos (BD) fue diseñada para almacenar y organizar la información necesaria según las especificaciones dadas por las especialistas para la construcción del sistema.

Los datos recopilados van a ser posteriormente procesados para su análisis. Pensando en ese siguiente paso y tomando en cuenta que los datos de tipo categórico requieren ser analizados numéricamente, la forma de almacenarlos será a través de códigos numéricos, cuyas claves se definen en los metadatos de los campos de la BD.

Como se trata de una base de datos de tipo relacional, la forma convencional de representarla es un diagrama entidad-relación. Como se ve en la Figura 5.1, el modelo está compuesto por 11 tablas, en las cuales se puede observar la información de los pacientes, el registro de sus respuestas a los cuestionarios y la puntuación obtenida de ellos.

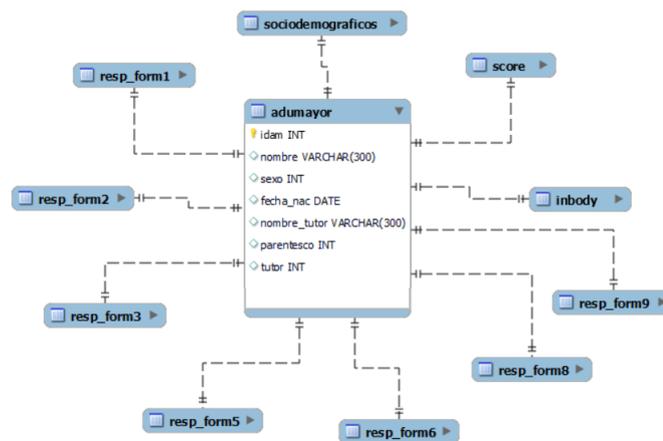


Figura 5.1: Diagrama entidad-relación

La tabla “adumayor” contiene la información básica del adulto mayor y en caso de necesitarlo, también de su tutor registrado.

Las tablas “sociodemograficos” e “inbody”, como sus nombres lo sugieren, guardarán la respectiva información del paciente. Mientras que la tabla “consentimiento” será la que confirme si el paciente firmó el documento donde da su consentimiento para participar en la investigación.

El resto de las tablas “resp_form” guardarán las respuestas que cada paciente proporcione en los cuestionarios preparados por las especialistas B. Finalmente, la tabla “score” registrará el resultado de todos los cuestionarios.

Adicionalmente, también se requirió contemplar el diseño de procedimientos almacenados y disparadores para facilitar la realización de algunos procesos, tales como la eliminación, actualización y alta de datos.

5. PLANEACIÓN DE LA CONSTRUCCIÓN DE LA HERRAMIENTA

Los procedimientos almacenados son procesos que se realizan en la BD al ser llamados por el usuario que adicionalmente puede proporcionar un o varios datos de entrada dependiendo de su función y de la misma forma, devolver o no datos de salida. Mientras que el disparador se activa al ocurrir un cambio en la BD y no depende directamente de la intervención del usuario.

El comportamiento de ambos procesos pueden verse descritos gráficamente en el diagrama 5.2.

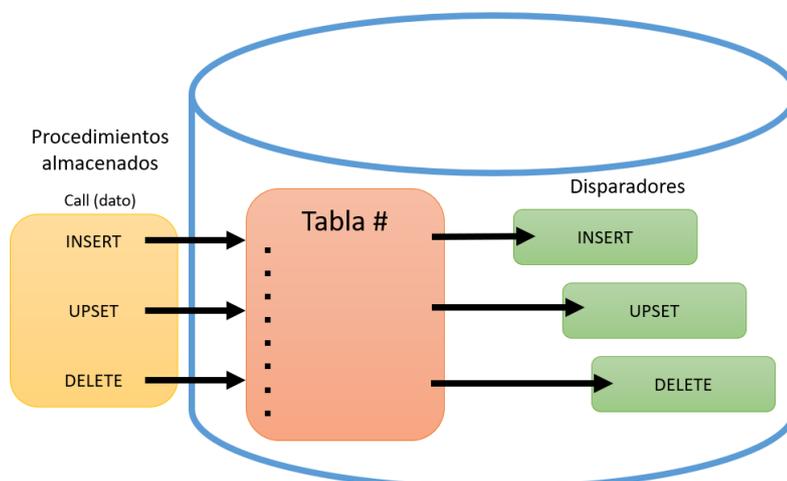


Figura 5.2: Diagrama de procesos almacenados y disparadores en una BD

5.3. Descripción de los procesos en el sistema

Partiendo de los requerimientos presentados en la sección anterior, se diseñó un sistema con los apartados necesarios para satisfacerlos. Entre los procesos que un especialista puede realizar con este sistema, están:

- Inicio de sesión: Es necesario que un especialista, como usuario del sistema, autentifique su identidad a través de un proceso en el que ingrese su nombre de usuario y contraseña. De lo contrario, la información almacenada en el sistema estaría expuesta a los peligros propios de un sistema en un entorno web.

- Ingresar datos sociodemográficos: Con este proceso se hace la recopilación de los datos de los adultos mayores que participen en la investigación, algunos de estos datos van a ser estudiados por las especialistas más adelante.
- Firma del consentimiento: Este proceso depende completamente de que se hayan ingresado los datos del adulto mayor. También tiene la intervención de otros dos actores además del especialista, los cuales serían el adulto mayor (Quién tiene que leer y firmar el documento de consentimiento) y, en caso de ser necesario, el tutor o responsable de esta persona (Si tuviera algún impedimento para consentir) también firme.
- Ingresar respuestas de cuestionarios: Un especialista puede saber a través de una serie de instrumentos de medición (cuestionarios) si un adulto mayor tiene las condiciones potenciales que propician la ideación suicida. Este proceso permite que los participantes en la investigación contesten estos cuestionarios y almacenar esas respuestas para posteriormente analizarlas.
- Ingresar datos Inbody: Esta investigación incluye una serie de pruebas físicas que se realizan con ayuda de un aparato denominado “Inbody”, así que también se requiere que se haga un registro de los datos recopilados con él.
- Editar información: Es común que se cometan errores en cualquier recopilación de información, si se diera el caso, es necesario que se le proporcione al especialista la opción de corregir o cambiar la información que se registró.
- Borrar información: Al igual que en el proceso anterior, hay muchas razones por las que pudiera ser necesario borrar la información registrada, así que esta misma opción debe estar presente si se requiere.
- Análisis estadístico de los datos: Los datos ingresados al sistema deben ser analizados por las especialistas una vez adquiridos, para ello es necesario aplicar

algunas técnicas de estadística básica como lo son las vistas en la Sección 2.3.

- **Análisis por Inteligencia Computacional:** Adicionalmente se integró una sección donde se pueden realizar análisis con algunos algoritmos de Inteligencia Computacional, que fueron mencionados en la Sección 2.2 con el fin de explorar el conocimiento dentro de los datos con diferentes técnicas y confirmar o rechazar las hipótesis de la investigación llevada a cabo por las especialistas.

El sistema debe tener diferentes vistas para que pueda interactuar con los actores que van a hacer uso o participarán de los procesos. Además de que, quien va a desarrollar el sistema, debe tener pleno entendimiento de cómo va a funcionar y los especialistas, que serían los usuarios finales, también deben dar su aprobación sobre el diseño.

Es por eso que dentro de las prácticas convencionales del desarrollo de software se acostumbra realizar diferentes diagramas que son útiles en la parte de planeación de las metodologías de trabajo. Tales diagramas servirán más adelante para el desarrollo, pues describen funcionalidades del sistema, diseño y la forma de interactuar con el usuario final. Aunado a eso, se le puede dar una representación sencilla a este usuario para darle una mejor idea del resultado final que obtendrá cuando el sistema esté terminado.

5.3.1. Diagrama Casos de uso

En el diagrama de casos de uso, se describen las secuencias de acciones que puede realizar el especialista al interactuar con el sistema. Así también el cómo participan otros actores externos en este proceso.

La Figura 5.3 presenta el diagrama generado para la creación del sistema propuesto.

Como se puede observar en la Figura, en este tipo de diagramas se puede apreciar principalmente la interacción que tienen los actores con el sistema. Por ejemplo, el especialista puede realizar varias acciones con el sistema como ingresar, editar, eliminar y analizar información. Sin embargo, también se puede ver que se requiere de la

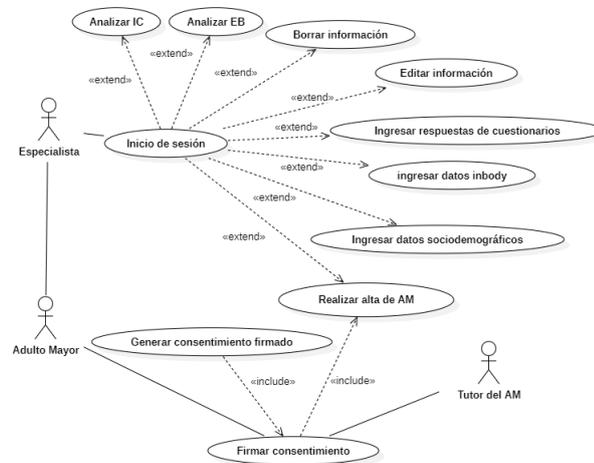


Figura 5.3: Diagrama de Casos de Uso del sistema a desarrollar (Elaboración propia)

participación de otros actores para que se realicen ciertos procesos.

El adulto mayor debe interactuar con el especialista para la recopilación de datos y firmar un documento donde consienta participar en la investigación, en caso de que este individuo lo amerite, también será necesaria la firma y consentimiento por parte de un tutor que pueda ayudarle en la toma de decisiones.

5.4. Wireframe

Para tener una vista preliminar de la página y tener una facilidad de diseño ágil, se crearon los wireframe de las secciones más importantes de la página.

Esto con el propósito de mostrarle a las especialistas una vista preliminar del sistema y adecuarlo a sus necesidades o gusto. Además de que fueron útiles como guía para el desarrollo del sistema.

5.4.1. Inicio de sesión

Es importante que el sistema cuente con un apartado para el inicio de sesión. De ese modo, el usuario puede identificarse, ingresar al sistema, como ya se explicó, esto es con el propósito de que ninguna persona ajena al trabajo de investigación de las

5. PLANEACIÓN DE LA CONSTRUCCIÓN DE LA HERRAMIENTA

especialistas pueda tener acceso a la información almacenada.

Como se ve en la Figura 5.4, para ingresar al sistema, el usuario solo necesita proporcionar su nombre de usuario y su contraseña.

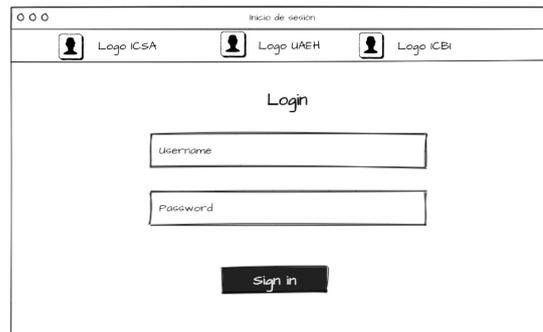


Figura 5.4: Wireframe del inicio de sesión del sistema (Elaboración propia)

5.4.2. Menú

En la vista del menú (ver Figura 5.5), el usuario puede ver y acceder a las opciones que tiene para ingresar información acerca de los adultos mayores que están participando en la investigación.

Por ejemplo, el ingreso de sus datos sociodemográficos, la firma del documento del consentimiento informado, la información del Inbody y los cuestionarios usados por las especialistas, entre otras opciones.

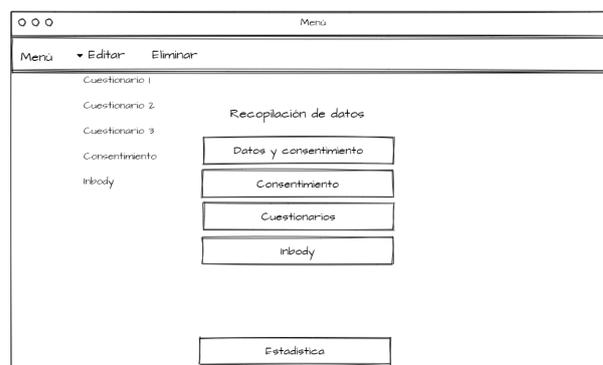


Figura 5.5: Wireframe del menú del sistema (Elaboración propia)

5.4.3. Alta de adulto mayor

El primer paso para comenzar a llenar todos los apartados con la información de los participantes en la investigación, sería tomar los datos más básicos del adulto mayor en cuestión (Nombre, sexo, edad, etc.) y los de su tutor (En caso de que lo requiera).

El nombre, identificador y otros datos serán útiles para ayudar a completar de una forma rápida y sin redundancias el resto de los apartados.

En la Figura 5.6 se muestra la forma que se desea para este apartado en particular.

El wireframe muestra un formulario web con el título "Datos sociodemográficos del adulto mayor". En la parte superior hay tres logos: "Logo ICSA", "Logo UAEH" y "Logo ICBI". El formulario contiene los siguientes campos:

Nombre	Apellido paterno	Apellido materno
Edad	Sexo	

Figura 5.6: Wireframe del formulario para ingresar los datos del adulto mayor (Elaboración propia)

5.4.4. Cuestionarios

Como es parte de la investigación de las especialistas, es necesario que los cuestionarios sean contestados por los propios adultos mayores. Por lo cual es preferible que el formato en el que estén dispuestos sea amigable con la vista y entendimiento de este tipo de personas y cuente con instrucciones en un lenguaje amable y conciso.

En la Figura 5.7 se puede apreciar que la idea es usar preguntas de opción múltiple que le permitan al adulto mayor expresar su respuesta simplemente con seleccionar la opción que deseen.

5. PLANEACIÓN DE LA CONSTRUCCIÓN DE LA HERRAMIENTA

The wireframe shows a window titled 'Cuestionario *'. At the top, there are three logos: 'Logo IC-SA', 'Logo UAFH', and 'Logo ICBI'. Below the logos, there is a section for 'Cuestionario para' followed by a dotted line. A message reads: 'Si no te sientes comodo(a) respondiendo alguna pregunta, hazle saber tus inquietudes al especialista que te está efectuando.' Below this is a section for 'Pregunta 1' followed by a dotted line. At the bottom, there are five radio buttons labeled '01', '02', '03', '04', and '05'. Under '01' is the word 'Nunca' and under '05' is the word 'Siempre'.

Figura 5.7: Wireframe de los cuestionarios (Elaboración propia)

5.4.5. Resultados

Para que las especialistas puedan ver los resultados de los cuestionarios es necesario presentarlos de manera ordenada junto al resto de las variables físicas que se tomaron en cuenta para el estudio.

En la Figura 5.8 se puede apreciar el formato en el que los resultados pueden ser analizados tomando en cuenta los valores de referencia proporcionados por las especialistas.

The wireframe shows a window with three logos at the top: 'Logo IC-SA', 'Logo UAFH', and 'Logo ICBI'. Below the logos are two input fields: 'Número del paciente' and 'No. Foto'. Below these is a table with three columns: 'Variables', 'Resultado', and 'Valores de referencia'. The table contains four rows of data.

Variables	Resultado	Valores de referencia
Variable 1	*	Min - Max
Variable 2	*	Min - Max
Variable 3	*	Min - Max
Variable 4	*	Min - Max

Figura 5.8: Wireframe de los resultados (Elaboración propia)

5.4.6. Registro de nuevos usuarios

Algunos pasantes van a participar en la investigación para apoyar a la recopilación de datos. Para que cada usuario del sistema tenga su propio acceso, es necesario que

exista un apartado en el cual se pueda hacer un registro de los usuarios.

Como se puede ver en la Figura 5.9, este apartado sería bastante similar al de Inicio de sesión, sin embargo contará con algunos rubros adicionales.

El wireframe muestra una interfaz de usuario para el registro de un nuevo usuario. En la parte superior, hay un título 'Registro de nuevo usuario' y tres botones de inicio de sesión con logotipos: 'Logo ICSA', 'Logo UAFH' y 'Logo ICEI'. El formulario principal, titulado 'Registro', contiene cuatro campos de entrada de texto: 'Nombre', 'Nombre de usuario', 'Tipo de usuario' y 'Contraseña'. Debajo de los campos, hay un botón 'Registrar'.

Figura 5.9: Wireframe de los registro de usuarios (Elaboración propia)

5.4.7. Tabla de estadísticas básicas

Las especialistas solicitaron la visualización de estadísticas básicas de las variables estudiadas en el sistema a través de una tabla para su análisis.

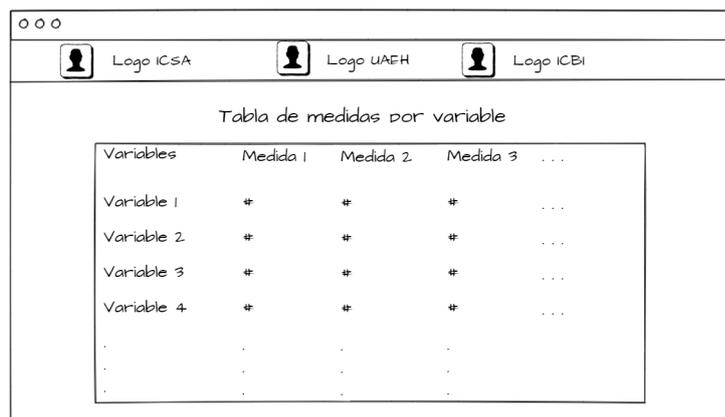
Se llegó a la conclusión de que en dicha tabla la información debería estar distribuida de manera que las variables quedaran enlistadas de forma vertical y las medidas estadísticas de forma horizontal para una correcta compresión y presentación, como se aprecia en la Figura 5.10.

5.4.8. Sección de análisis de datos

Para el apartado en el que se van a analizar los datos, fue solicitada la inclusión hay tres elementos clave: la selección de las variables a analizar, una tabla donde se muestren los datos del análisis realizado y la presentación de los datos a través de herramientas de visualización.

En la Figura 5.11 se presenta la forma en la que estos elementos estarían distribuidos en la vista del sistema.

5. PLANEACIÓN DE LA CONSTRUCCIÓN DE LA HERRAMIENTA



The wireframe shows a window with three user profile icons at the top: Logo ICSA, Logo UAFH, and Logo ICBI. Below them is the title "Tabla de medidas por variable". The main content is a table with the following structure:

Variables	Medida 1	Medida 2	Medida 3	...
Variable 1	#	#	#	...
Variable 2	#	#	#	...
Variable 3	#	#	#	...
Variable 4	#	#	#	...
.
.
.

Figura 5.10: Wireframe de la tabla de estadísticas básicas

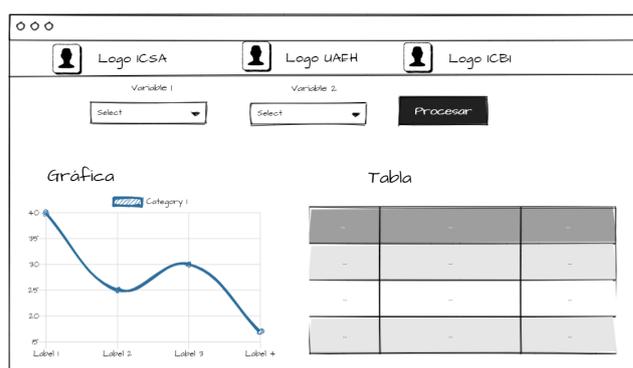


Figura 5.11: Wireframe de la sección de análisis de datos

Desarrollo de la Herramienta Propuesta

Durante esta segunda fase de la metodología, se llevaron las actividades necesarias para cumplir con los requerimientos que las especialistas necesitan que contenga el sistema. Esto será usando como guía los diagramas realizados en el capítulo anterior.

6.1. Implementación de un servidor web local para las pruebas

Debido a la necesidad de que el sistema se encuentre disponible para diversas plataformas y múltiples usuarios y, además, de que la información almacenada se encuentre actualizada en tiempo real, se propone desarrollar una aplicación web como solución ideal para satisfacer dichas necesidades.

Para empezar, es necesario implementar un servidor web local que sea capaz de recibir peticiones y devolver respuestas a un usuario para comenzar las pruebas del sistema de forma sencilla y rápida para después subirlo a un servidor web más robusto. En esta investigación en particular, como se requirió hacer una unión entre varias herramientas de programación y desarrollo web (Python, JavaScript, HTML, etc), se hizo uso de un framework conocido como Flask (cuyas cualidades ya se explicaron en el [Capítulo 2](#)) para realizar esta actividad y obtener el entorno propicio para comenzar el desarrollo.

6. DESARROLLO DE LA HERRAMIENTA PROPUESTA

Para comenzar con la instalación de Flask es necesario tener instalado Python, Pip (un administrador de paquetes de archivos escritos en Python) y una carpeta en la cual se van a guardar los archivos del proyecto. Este proceso se puede visualizar en la Figura 6.1.

```
(venv) PS C:\Users\ \Desktop\rdsam> Python --version
Python 3.11.0
(venv) PS C:\Users\ \Desktop\rdsam> pip --version
pip 23.2.1 from C:\Users\ \Desktop\rdsam\venv\Lib\site-packages\pip
(venv) PS C:\Users\ \Desktop\rdsam>
```

Figura 6.1: Carpeta con Python y Pip instalados

A continuación, se debe crear un entorno virtual, que son carpetas que permiten guardar los paquetes de un proyecto para que no se mezclen con otros y tener una mejor organización al momento de desarrollar.

Haciendo uso de Pip, ya se puede instalar el módulo de Flask en el entorno virtual para implementar el servidor. Para que Flask realice la función deseada, es necesario crear un archivo a través del cual se importe el módulo y al ejecutarlo, empezará a funcionar el servidor, que por default ejecuta el servidor en el puerto 5000 como se ve en la Figura 6.2.

```
(venv) PS C:\Users\ \Desktop\rdsam> python app.py
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in production.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 649-675-561
```

Figura 6.2: Ejecución del servidor

6.2. Construcción de la base de datos

Una base de datos es indispensable para casi cualquier sistema que requiera el almacenamiento de datos y/o la consulta de los mismos. Para construir la de este sistema en particular, se hizo uso de un software para el manejo de bases de datos

conocido como PostgreSQL. Al haber hecho un apropiado diseño de la BD, es poco probable que exista algún inconveniente al momento de construirla.

En este caso en específico, el único obstáculo que se presentó para llevar a cabo esta actividad, fueron las diferencias gramaticales que hay entre el software usado (PostgreSQL) y el lenguaje SQL convencional, ya que, algunas instrucciones no podían ser interpretadas de la misma manera y se tuvo que hacer una investigación sobre la traducción de las palabras reservadas.

Las pruebas a la conexión de la BD y las relaciones entre sus tablas fueron satisfactorias, debido a que fue implementada de acuerdo a el diagrama mostrado en la Figura 5.1.

6.3. Desarrollo de la aplicación web

El desarrollo de aplicaciones web es la implementación de técnicas enfocadas a este campo (como JavaScript, HTML, Bootstrap, etc.), para la creación de software al cual se puede acceder desde un navegador web o dispositivo móvil que cuente con conexión a una red de internet.

Para el sistema realizado en este trabajo, se integraron varias de las técnicas mencionadas en el párrafo anterior con el propósito de que se complementen entre sí y dar por resultado un sistema con la calidad necesaria para cumplir con los propósitos para los cuales se requiere desarrollarlo.

Una vez que el servidor se encontró en un estado funcional, se hizo uso del Lenguaje de Marcado de Hipertexto (HTML), el cual es básico para estructurar el contenido de una página web. En conjunto con librerías como Bootstrap, se logró generar páginas visualmente más agradables en menos tiempo.

La mayoría de los procesos que involucraron a la base de datos, como el alta de datos, la visualización, edición y eliminación de los mismos, fueron programados con

6. DESARROLLO DE LA HERRAMIENTA PROPUESTA

Python a través de Flask con el uso de los módulos disponibles en este lenguaje para conectarse con PostgreSQL.

Sin embargo, para conseguir que el sistema y el usuario interactúen, también se hizo uso de un lenguaje de programación especializado en contenido web, JavaScript.

6.3.1. Inicio de sesión

Como primer paso para realizar cualquier proceso en el sistema, es necesario que los usuarios se identifiquen con su respectivo nombre de usuario y contraseña.

La interfaz para realizar este paso (Ver Figura 6.3) contiene elementos visuales como los logotipos de la Universidad Autónoma del Estado de Hidalgo, el Instituto de Ciencias Básicas e Ingeniería (ICBI) y el Instituto de Ciencias de la Salud (ICSA), que colaboraron para la realización de este sistema, con el propósito de que cuando los pacientes interactúen con el sistema, tengan conocimiento de que participan en una investigación realizada por estudiantes que pertenecen a estas entidades.

En la parte central hay un formulario en el que se solicitan los datos mencionados anteriormente y un botón que al presionarlo da acceso al sistema si el usuario y la contraseña se encuentran registrados en la BD.

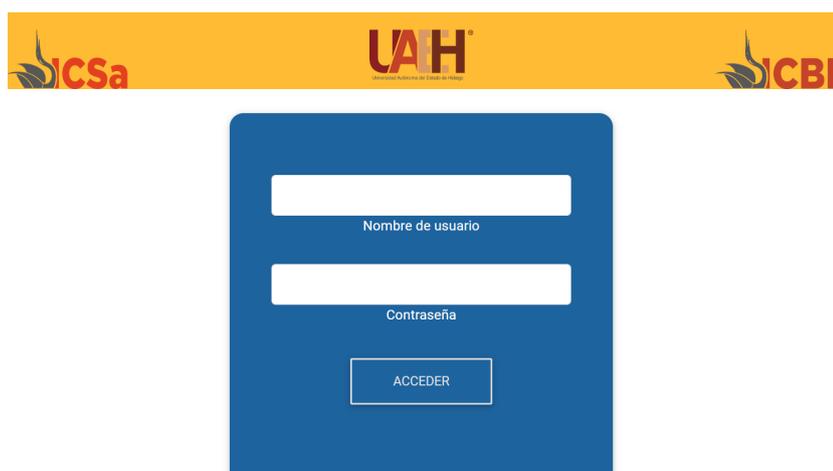
La imagen muestra la interfaz de inicio de sesión del sistema. En la parte superior hay una barra amarilla con los logotipos de ICSa, UAH (Universidad Autónoma del Estado de Hidalgo) y ICBI. Debajo de esta barra, sobre un fondo azul, se encuentra un formulario con dos campos de texto blancos: el superior está etiquetado como 'Nombre de usuario' y el inferior como 'Contraseña'. Debajo de los campos hay un botón rectangular con el texto 'ACCEDER'.

Figura 6.3: Interfaz para el inicio de sesión del sistema

Pensando en que las especialistas reciban ayuda de pasantes con menos experiencia con el manejo de datos, se crearon dos tipos de usuarios, uno para los administradores (especialistas encargadas de la investigación) y otro para los pasantes que las apoyen. Se crean estos roles con el fin de restringir el acceso a algunos procesos a los usuarios con menos experiencia y grado de responsabilidad sobre los datos recopilados.

6.3.2. Menú principal

Dentro del menú principal, del lado superior derecho se encuentra un apartado en el cual aparece el nombre del usuario y una etiqueta, que sirve para cerrar la sesión.

El menú principal para los administradores es el que se muestra en la Figura 6.4. En este menú podrán acceder a los diferentes procesos que ejecuta el sistema. Además, en la parte superior izquierda se tiene un menú donde el usuario puede editar la información de los pacientes, o en algún caso en específico, borrar el registro de alguno de ellos.

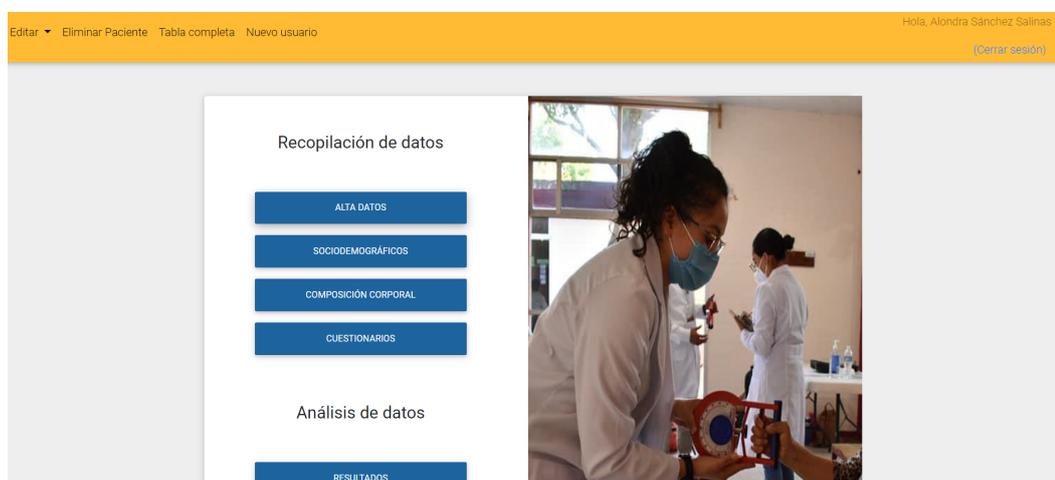


Figura 6.4: Menú principal para administradores

Como se mencionó anteriormente, los pasantes no podrán realizar algunos procesos. En la Figura 6.5 se puede ver que ellos no contarán con los íconos de Editar y Eliminar paciente, ya que estos procesos sólo se llevan a cabo si alguna especialista así lo determina.



Figura 6.5: Menú principal para pasantes

Las acciones que puede realizar el usuario en el sistema, se explican a detalle a continuación.

6.3.3. Alta de datos y consentimiento informado

Para comenzar con el registro de un adulto mayor, lo primero que se debe hacer es registrar sus datos principales dentro del formulario que se puede apreciar en la Figura 6.6.

Si está acompañado de algún familiar o tutor que vaya a ayudarlo a realizar su participación en la investigación, también deberá registrarse en esta interfaz.

Una vez llenado el formulario, la información es registrada y con ello se puede pasar directamente al siguiente proceso en donde el paciente lea y firme el documento del consentimiento informado, donde se le explica detalladamente en que consta la investigación, su propósito, la privacidad de su información, etc.

Si una vez leído el documento, el adulto mayor acepta participar, es necesario que firme de forma digital en el apartado que se observa en la Figura 6.7. En caso de que se haya registrado un tutor, también deberá firmar del lado derecho debajo de su nombre.

Una vez firmado en este elemento, se puede generar el archivo PDF donde queda

Formato de identificación

(Este formulario debe ser llenado exclusivamente por el personal de campo, en ningún momento debe llenarlo el participante o su acompañante)

Nombre completo del adulto mayor

Ej. José Pérez Martínez

Sexo

Elija el sexo

Fecha de nacimiento

(Solicite INE u otra identificación oficial para verificar)

1963-01-01

Figura 6.6: Formulario para el alta de datos de un paciente

registrado que el adulto mayor aceptó formar parte de la investigación y se genera el archivo PDF como se ve en la Figura 6.8.

6.3.4. Recolección de datos faltantes

Para llevar a cabo esta acción, se hace uso de un conjunto de interfaces que a través de tablas solicitan los datos de aquellos pacientes de los cuales no se ha recopilado la información de cada apartado, como la obtenida del Inbody, los cuestionarios o los datos sociodemográficos. Estas tablas son útiles para su recolección, evitando que haya problemas al momento de recordar cuáles son aquellos pacientes que faltan de registrar en los diferentes procesos.

La tabla contiene tres columnas en las cuales se muestra el folio del paciente, su nombre y la acción que va a realizarse, tal como se muestra en la Figura 6.9.

6.3.5. Datos Sociodemográficos

Los datos sociodemográficos son información sobre las condiciones sociales, económicas, demográficas y geográficas en las que vive el paciente y se relacionan con su calidad de vida. Estos datos pueden ayudar a revelar factores protectores y de riesgo ante la

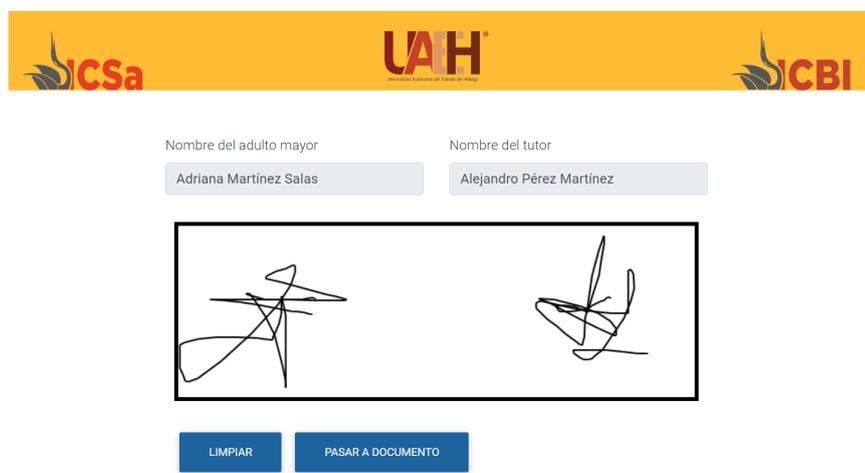


Figura 6.7: Apartado para firmar digitalmente el consentimiento informado

ideación suicida de un adulto mayor y es por eso que es importante recolectarlos.

Como se ve en la Figura 6.10, la interfaz en la cual se van a registrar estos datos, cuenta con dos campos previos al formulario en los que se corrobora el nombre del paciente y su número de folio para que los especialistas tengan presente en todo momento a que paciente están atendiendo.

Finalmente se encuentran los campos que, como se aclara en la etiqueta debajo del título de la interfaz, debe ser llenado únicamente por los especialistas.

6.3.6. InBody

Las pruebas InBody consisten en hacer una evaluación de las funciones físicas del cuerpo de una persona (en este caso, los adultos mayores) con ayuda de un instrumento diseñado para dicha tarea.

Entre los datos que se pueden recopilar con esta tecnología, se encuentra la fuerza muscular, velocidad de la marcha, índice de masa corporal, entre otras.

En la Figura 6.11 se puede observar que la estructura es parecida a la vista en los datos sociodemográficos.

Adicionalmente, los campos de este formulario están condicionados para no aceptar

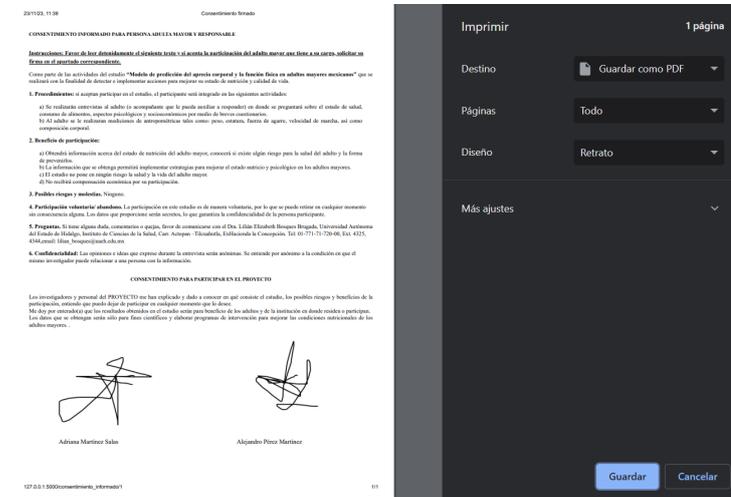


Figura 6.8: Generación del archivo PDF

caracteres que no sean números, debido a que los datos que se van a registrar, son de tipo numérico.

6.3.7. Cuestionarios

Los cuestionarios proporcionados por las especialistas, preferentemente, deben ser contestados por los propios adultos mayores a menos que alguna condición especial se los impida y requieran de la asistencia de un tercero.

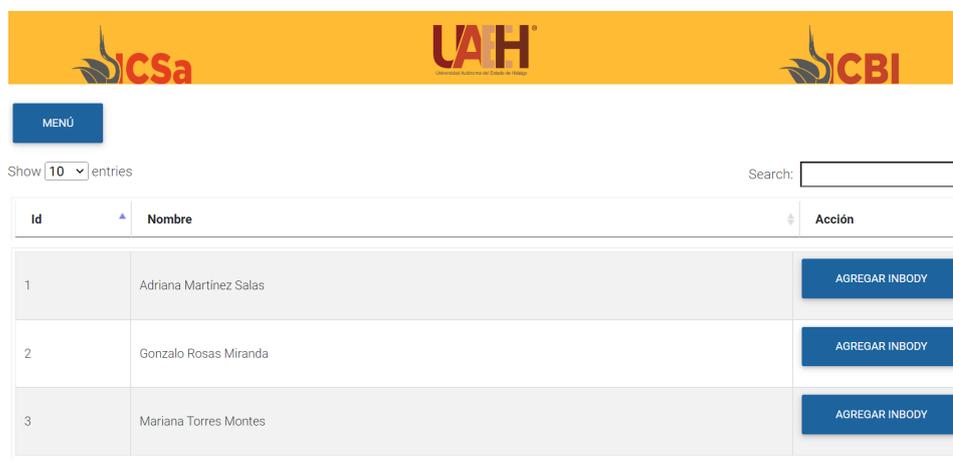
Es por lo anterior que, como se puede ver en la Figura 6.12, el formato en el que se presentan estas preguntas está dispuesto de manera en la que una persona de edad avanzada pueda contestar haciendo el menor esfuerzo posible.

Estos cuestionarios están organizados para ser contestados uno tras otro hasta completarlos todos y adicionalmente, también se programaron para que todas las preguntas sean respondidas antes de pasar al siguiente cuestionario.

6.3.8. Editar y eliminar

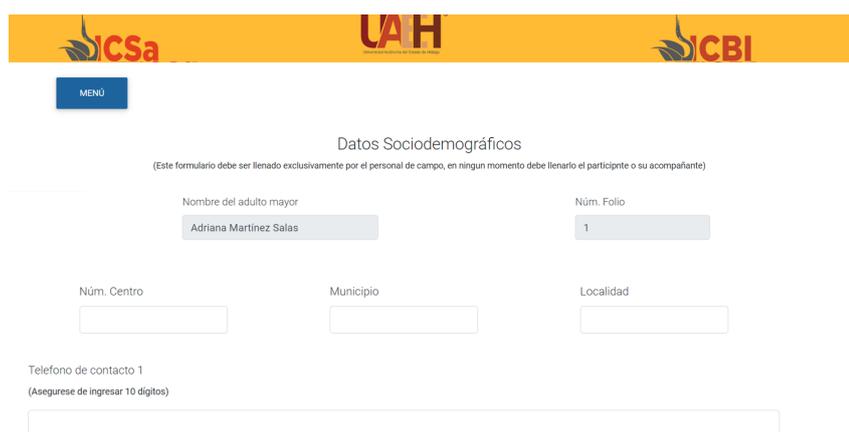
El acceso a estos procesos se encuentra en el encabezado del sitio web y está disponible únicamente para los administradores.

6. DESARROLLO DE LA HERRAMIENTA PROPUESTA



Id	Nombre	Acción
1	Adriana Martínez Salas	AGREGAR INBODY
2	Gonzalo Rosas Miranda	AGREGAR INBODY
3	Mariana Torres Montes	AGREGAR INBODY

Figura 6.9: Tabla en el sistema



Datos Sociodemográficos
(Este formulario debe ser llenado exclusivamente por el personal de campo, en ningún momento debe llenarlo el participante o su acompañante)

Nombre del adulto mayor: Adriana Martínez Salas

Núm. Folio: 1

Núm. Centro:

Municipio:

Localidad:

Telefono de contacto 1
(Asegurese de ingresar 10 dígitos):

Figura 6.10: Interfaz para recolectar los datos sociodemográficos de los pacientes

Para editar la información proporcionada por los adultos mayores, es necesario elegir cual es el formulario que se va a editar tal y como se aprecia en la Figura 6.13.

Eliminar a un paciente borrará toda la información registrada en la base de datos.

Ya que ambos procesos comprometen la información de la BD, su uso es exclusivo para usuarios de tipo administrador.

Figura 6.11: Interfaz para recolectar los datos de la prueba InBody de los pacientes

Figura 6.12: Cuestionario a contestar por el adulto mayor

6.3.9. Resultados

Los resultados de los cuestionarios deben ser analizados por un especialista en el área de salud mental para poder dar un diagnóstico confiable debido a que se trata de variables complejas. Con el fin de ayudar a las especialistas a ver los resultados, se diseñó una tabla, como la que se puede ver en la Figura 6.14, en la que se puso a su alcance el nombre de la variable estudiada, el resultado obtenido y los límites de referencia para que puedan deliberar si el paciente está en riesgo o no de cometer suicidio.

6. DESARROLLO DE LA HERRAMIENTA PROPUESTA



Figura 6.13: Menú desplegado de opciones para editar y eliminar un registro

Una interfaz de usuario que muestra los resultados de un paciente. Incluye un botón "CANCELAR", campos de texto para "Nombre del adulto mayor" (Miguel Bartolo Miranda) y "Núm. Folio" (9). Abajo hay una tabla con tres columnas: Variable, Resultado y Valores de referencia.

Variable	Resultado	Valores de referencia
Apreciación corporal	65	13 - 65 (Las puntuaciones más altas indican una mayor apreciación del cuerpo)
Gratitud	30	6 - 42 (Las puntuaciones más altas indican una mayor gratitud)
Autoestima	22	<15 Bajo 15 - 25 Normal >30 Alto
Depresión	21	0 - 21 (Las puntuaciones más altas indican una mayor depresión)

Figura 6.14: Tabla en la que se muestran los resultados del paciente y los valores de referencia

6.3.10. Registro de nuevos usuarios

En esta sección, un usuario ya registrado puede dar de alta a nuevos usuarios. Como se ve en la Figura 6.15, este proceso consiste en proporcionar información básica del usuario para el sistema y con ello ya podría iniciar sesión desde otro dispositivo.

Cabe destacar que a esta sección solo podrán acceder los usuarios con un rol de tipo Administrador, con el fin de que haya un control sobre las personas que son registradas como usuarios del sistema.

Figura 6.15: Sección en la que se registran los nuevos usuarios

6.3.11. Tabla de estadísticas básicas

Con la finalidad de contar con una herramienta que permita en un solo espacio aplicar métricas de estadística básica, se propone en este sistema, desarrollar un apartado que incluya algunas de estas.

De cada variable en el sistema se podrán obtener los valores de media, desviación estándar, máximo valor, mínimo valor, entre otras, se presentan en el sistema de forma ordenada en una tabla como se observa en la Figura 6.16.

Variable	Media	Desviación estandar	Mínimo	Máximo	Percentil 1	Percentil 2	Percentil 3
sexo	0.3717948717948718	0.48641210572593335	0.0	1.0	0.0	0.0	1.0
tutor	1.8074923076923077	0.3966644140109588	1.0	2.0	2.0	2.0	2.0
parentesco	0.8589743589743589	1.7412069191465861	0.0	5.0	0.0	0.0	0.75
num_centro	2254.0897435897436	9647.071417787549	0.0	43993.0	0.0	0.0	0.0
div	3.2435897435897436	1.8669086708642135	1.0	6.0	2.0	2.0	6.0
mgro	1.3974358974358974	0.7951107672526716	1.0	5.0	1.0	1.0	2.0

Figura 6.16: Sección en la que se presenta la tabla de estadísticas básicas

De esa manera, pueden analizar directamente los resultados de las medidas de cada variable sin necesidad de escribir código haciendo el menor esfuerzo posible.

6.3.12. Sección de análisis de datos

Dentro de este apartado se pueden hacer varios tipos de análisis de datos dentro de los cuales se incluyen los mencionados en las Secciones 2.3 y 2.2.

Para realizar estos estudios, el primer paso que se debe seguir es seleccionar las variables a analizar en la parte superior de la página y presionar el botón que ejecutará el proceso de los datos y entonces aparecerán los valores deseados en la tabla ubicada del lado inferior derecho y la gráfica correspondiente del lado inferior izquierdo. Este proceso por parte del análisis estadístico se puede apreciar en la Figura 6.17

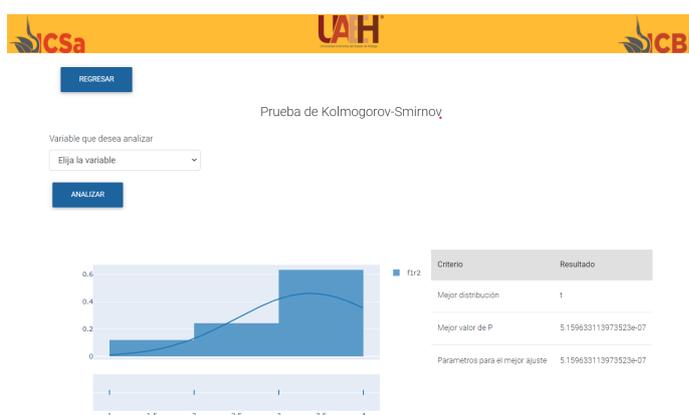


Figura 6.17: Sección en la que se realiza la prueba de Kolmogorov-Smirnov

Para ejemplificar la forma en la que el sistema hace un análisis usando Inteligencia Computacional en la Figura 6.18 se puede ver que para usar el algoritmo K-Means se obtiene la gráfica previa que indica cual podría ser el mejor número de grupos.

A continuación, en la Figura 6.19 se observa como un especialista puede ingresar el número de grupos con los que desee experimentar ya sean los recomendados por la gráfica anterior o los que requiera por decisión propia con un tope de hasta siete grupos. Finalmente se obtiene la gráfica de tales grupos y la tabla con la caracterización de los



Figura 6.18: Sección en la que se realiza el agrupamiento de los datos con K-Means

centroides de cada grupo.



Figura 6.19: Sección en la que se aprecia la gráfica de grupos generados por K-Means y la tabla de caracterización de los centroides

Además de K-Means, el sistema también puede implementar algoritmos de regresión lineal y logística para obtener modelo de los datos.

6.4. Instalación del sistema en un servidor en la nube

Para que el sistema pueda ser usado en cualquier momento y en cualquier lugar, es necesario que esté alojado en un servidor que funcione todo el tiempo. De manera local esto no es posible, es por ello que se eligió la plataforma Microsoft Azure para obtener

6. DESARROLLO DE LA HERRAMIENTA PROPUESTA

tal servicio.

Además de ser una de las plataformas más confiables, variadas, completas y que cuenta con un apoyo para estudiantes que ofrece un saldo de 100 dolares para usar sus servicios, también es compatible con todas las herramientas utilizadas para el desarrollo del sistema.

6.4.1. Creación del servidor de base de datos

Para crear un servidor de base de datos en Azure, primero se debe escoger cual será el manejador de base de datos que se va a utilizar, para este trabajo se eligió el servidor flexible de Azure Database for PostgreSQL.

A continuación, se deben llenar los campos que Azure irá mostrando para configurar el servidor como se muestra en la Figura 6.20.

Inicio > Crear un recurso > Marketplace > Azure Database for PostgreSQL Flexible Server >
Servidor flexible de Azure Database for PostgreSQL

Microsoft

⚠ No se pueden cambiar los nombres de servidor, los métodos de conectividad de red y la redundancia de copia de seguridad después de crear el servidor. Revise estas opciones cuidadosamente antes de realizar el aprovisionamiento.

Detalles del proyecto
Seleccione la suscripción para administrar recursos implementados y los costos. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción *
Grupo de recursos *
[Crear nuevo](#)

Detalles del servidor
Especifique la configuración necesaria para este servidor, incluida la selección de una ubicación y la configuración de los recursos de proceso y almacenamiento.

Nombre del servidor *
Región *
Versión de PostgreSQL *
Tipo de carga de trabajo * Desarrollo Producción
Proceso y almacenamiento *

Costes estimados

- SKU de proceso: MXN 5483.47/mes
Standard_D4ads_v5 (4 núcleos virtuales, 4 MXN 1370.87 por núcleo virtual)
- Almacenamiento: MXN 310.66/mes
Almacenamiento seleccionado: 128 x 128 GiB (MXN2.43 por GiB) 2.43
- Alta disponibilidad: MXN 5794.13/mes
Alta disponibilidad con redundancia de zona
- Ancho de banda
En el caso de la transferencia de datos saliente entre servicios de ...

[Revisar y crear](#) [Siguiente: Redes >](#)

Figura 6.20: Vista de Azure para la configuración del servicio del servidor de base de datos

Se debe tener especial cuidado al llenar estos campos y asegurarse de que la configuración del servidor sea compatible con las necesidades del sistema. sin embargo, la plataforma ofrece la opción de cambiar la configuración si se llegase a cometer un error.

6.4.2. Creación del servidor web en la nube

Al igual que el servidor de base de datos, el servidor web también debe ser configurado al momento de ser creado, aunque corregir un error en la configuración de este servicio puede ser problemático. Por ejemplo, si se escribe mal el nombre del dominio puede causar varios problemas al subir los archivos del sistema o al compartir el URL asignado con los usuarios.

En la Figura 6.21 se presenta la vista con la que se comienza la configuración del servidor.

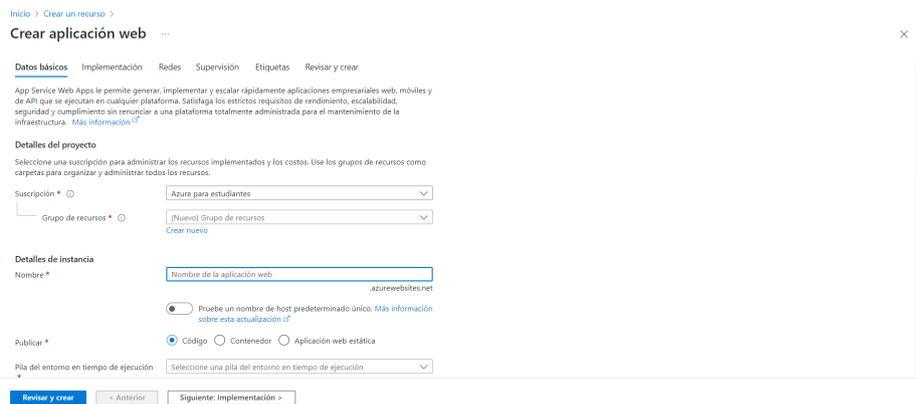


Figura 6.21: Vista de Azure para la configuración del servicio del servidor web

6.4.3. Instalación del sistema en el servidor web

Como la IDE utilizada usada para el desarrollo del sistema (VS Code) y el proveedor de servicios en la nube (Azure) fueron desarrollados por la misma corporación, hay extensiones que facilitan la conexión entre ellos y hacen que subir los proyectos a producción sea posible en un tiempo muy corto y en pocos pasos.

Solo hace falta instalar la extensión de Azure en VS Code y se agrega una función al menú extendido de las carpetas con el cual se puede desplegar el proyecto en el servidor que se creó como se muestra en la Figura 6.22.

6. DESARROLLO DE LA HERRAMIENTA PROPUESTA

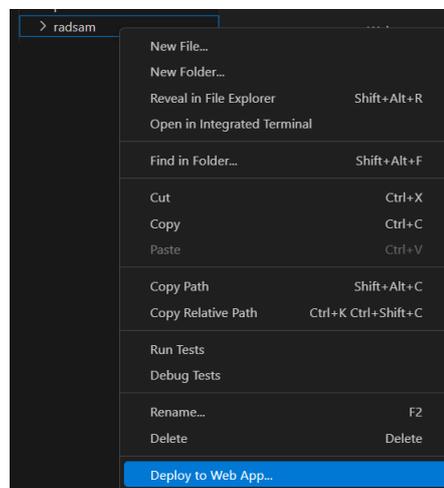


Figura 6.22: Vista de Azure para la configuración del servicio del servidor de base de datos

Una vez realizado el paso anterior solo hace falta esperar a que se despliegue el sistema y cuando finalice esta acción, en la consola de la IDE aparecerá un mensaje que informará si el despliegue se realizó con éxito o no como el que aparece en la Figura 6.23.

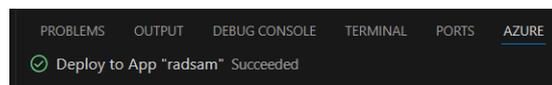


Figura 6.23: Vista de Azure para la configuración del servicio del servidor de base de datos

Validación de la Herramienta

Es importante contar con una herramienta que ayude a evaluar un software, de esa forma es posible garantizar su calidad, su adecuación a las necesidades del usuario y su capacidad para cumplir con los objetivos establecidos para tal sistema. Por ello es que en este capítulo se describen las herramientas usadas para este propósito y los resultados obtenidos.

7.1. Evaluación de requerimientos

Como parte de la validación del sistema, una de las métricas usadas fue la evaluación de requerimientos por parte de las especialistas.

Una evaluación de requerimientos se realiza cuando el usuario final del software a evaluar (las especialistas), da una puntuación en una escala determinada sobre si el sistema cumple con los requerimientos solicitados inicialmente. De ser así, se otorga la puntuación máxima y se considera que se ha cumplido con el requerimiento, de lo contrario, es necesario que el usuario final proporcione una retroalimentación para mejorar el apartado que debería cumplir con el requerimiento o rediseñarlo. Esto último depende de que tan baja sea la puntuación dada por el usuario final.

Para evaluar este sistema solo se tomó en cuenta la evaluación de las dos especialistas principales con quienes se realizó la planeación del desarrollo del sistema.

7. VALIDACIÓN DE LA HERRAMIENTA

Se les pidió que dieran una calificación del 1 al 5 donde 1 significaba que no se había cumplido con el requerimiento y 5 que el requerimiento se había cumplido tal y como se solicitó.

Como se puede apreciar en la Tabla 7.1, todos los requerimientos se cumplieron, por lo que el sistema cubrió todas los rubros solicitados.

Requerimiento	especialista_1	especialista_2
Portabilidad entre diferentes plataformas. Es indispensable el uso en equipo de cómputo y dispositivos móviles.	5	5
Actualización de los datos en tiempo real.	5	5
Acceso de múltiples usuarios a la vez desde diferentes lugares.	5	5
Adquisición de datos de forma independiente, de acuerdo al tipo de información del adulto mayor, para agilizar el proceso.	5	5
Interfaces responsivas de los formularios con la finalidad de que sean amigables con la vista de un adulto mayor promedio.	5	5
Apartado para firmar digitalmente un documento de “Conocimiento informado”.	5	5
Disponibilidad de poder ver la información de un adulto mayor en ciertos apartados.	5	5
Registro adecuado de los valores de las respuestas a las preguntas directas e inversas.	5	5

Tabla 7.1: Tabla de la evaluación de requerimientos por parte de las especialistas

7.2. Evaluación de Sistema de Escalas de Usabilidad

El Sistema de Escalas de Usabilidad (SUS por sus siglas en ingles) es una metodología desarrollada por Brooke (78) utilizada para medir y evaluar la usabilidad de sistemas interactivos. Este enfoque se basa en la utilización de escalas y métricas específicas para evaluar diferentes aspectos de la usabilidad, como la eficacia, la eficiencia, la satisfacción del usuario y otros aspectos relacionados con la experiencia del usuario.

La evaluación consta de 10 enunciados:

1. Creo que me gustaría utilizar este sistema con frecuencia
2. Encontré el sistema innecesariamente complejo
3. Pienso que el sistema es fácil de usar
4. Creo que necesitaría el apoyo de un técnico para poder utilizar este sistema
5. Encontré que las diversas funciones de este sistema estaban bien integradas
6. Pienso que hay demasiada inconsistencia en este sistema
7. Me imagino que la mayoría de la gente aprendería a utilizar este sistema muy rápidamente
8. Encontré el sistema muy complicado de usar
9. Me sentí muy seguro usando el sistema
10. Tuve que aprender muchas cosas antes para poder empezar con este sistema

Esta evaluación se contesta siguiendo la escala de Likert, que solo ofrece 5 opciones:

- Totalmente en desacuerdo
- En desacuerdo
- Neutro
- De acuerdo
- Totalmente de acuerdo

Como se puede observar, los enunciados impares son afirmaciones positivas hacia el sistema y los pares son negativas. Para obtener el resultado de la evaluación se utiliza la formula:

7. VALIDACIÓN DE LA HERRAMIENTA

$$resultado = [(\sum_{i=0}^4 e_{2i+1}) + 20 - (\sum_{i=0}^5 e_{2i})]2.5 \quad (7.1)$$

donde e es la respuesta del enunciado y n el número del enunciado.

El sistema fue usado por las dos especialistas con quienes se colaboró para la realización de este sistema y cinco pasantes calificados para realizar la recopilación de datos, por lo que se obtuvieron siete evaluaciones cuyos resultados se muestran en la [Tabla 7.2](#).

Enunciado	esp_1	esp_2	esp_3	esp_4	esp_5	esp_6	esp_7
Creo que me gustaría utilizar este sistema con frecuencia	5	5	4	4	5	5	4
Encontré el sistema innecesariamente complejo	1	2	2	2	4	2	4
Pienso que el sistema es fácil de usar	5	4	5	4	3	5	4
Creo que necesitaría el apoyo de un técnico para poder utilizar este sistema	1	3	1	2	4	1	2
Encontré que las diversas funciones de este sistema estaban bien integradas	5	4	5	4	4	5	4
Pienso que hay demasiada inconsistencia en este sistema	1	2	2	2	4	2	2
Me imagino que la mayoría de la gente aprendería a utilizar este sistema muy rápidamente	5	5	4	4	1	5	3
Encontré el sistema muy complicado de usar	1	2	2	1	4	1	2
Me sentí muy seguro usando el sistema	5	5	4	5	3	5	4
Tuve que aprender muchas cosas antes para poder empezar con este sistema	2	2	3	4	2	2	1
Resultado	97.5	80	80	75	45	92.5	70
Promedio	77.14						

Tabla 7.2: Tabla de las evaluaciones SUS de los especialistas que usaron el sistema

Jeff Sauro en su investigación donde mide la usabilidad de SUS (79), dice que el promedio de los sistemas en su estudio es de 68 sobre 100. Un resultado por debajo de esa cifra indica que hay varios aspectos por corregir.

Ya que, el promedio de las evaluaciones de los especialistas fue de 77.14, se puede decir que está por encima del promedio y entra en una categoría aceptable.

Conclusiones

En México hay pocos organismos que se encargan de prevenir el suicidio en la población. Y ya que el capital humano es el recurso más valioso en las sociedades, es importante prevenir que las personas tomen la decisión de terminar con su propia vida.

Según el Consejo Nacional de Población (CONAPO) en México en el 2010, el 8.8% de la población estaba representada por adultos mayores. Para el 2050 se estima que la población geriátrica crecerá hasta ser el 28% de la población (80), es por ello que es importante cuidar de la calidad de vida de los adultos mayores.

Con la finalidad de apoyar a los especialistas en el diagnóstico del riesgo de que un adulto mayor cometa suicidio, se desarrolló un sistema agilizó los procesos de recolección y organización de datos.

Estos datos también pueden ser analizados por los expertos desde el mismo sistema ya que cuenta con algunas funciones que permiten hacer un análisis estadístico e implementar algunos algoritmos de Inteligencia Computacional.

Con esta información se pretende encontrar aquellos factores que más impacto tienen en el pensamiento negativo hacia la propia vida en poblaciones geriátricas, factores que pueden ser de tipo psicológicos, sociales, físicos.

7.2.1. Aportaciones

Dentro de las aportaciones que se realizaron con la realización de este trabajo se encuentran:

7. VALIDACIÓN DE LA HERRAMIENTA

- El sistema que permite administrar la información refrenté a adultos mayores y a los factores que inciden en que un adulto mayor esté en riesgo de poner en riesgo su vida intencionalmente.
- La automatización en el proceso de adquisición de datos por los especialistas.
- Inclusión de técnicas de análisis de datos que permiten generar modelos de datos que apoyan el trabajo de las especialistas en el diagnóstico del riesgo de los pacientes a cometer suicidio.

7.2.2. Trabajo futuro

Para dar continuación al trabajo realizado hasta el momento, se pretende realizar:

- Incorporación de otras técnicas de predicción y clasificación de datos.
- Generar modelos que permitan emitir recomendaciones para la prevención del suicidio en la población geriátrica.

Referencias

- [1] Organización Mundial de la Salud (OMS), “Cie-11 para estadísticas de mortalidad y morbilidad.” <https://icd.who.int/browse11/1-m/es>, 2018.
- [2] E. S. Shneidman, *Definition of suicide*. New York: John Wiley Sons, 1985.
- [3] G. Barrientos de Llanos and A. Clavijo Portieles, *Conducta suicidad y su manejo en atención primaria de salud*, ch. 106. Ecimed, 2008.
- [4] M. R. A. M. E. G. C. A. Ribot Reyes, Victoria de la Caridad Alfonso Romero, “Suicidio en el adulto mayor,” *Revista Habanera de Ciencias Médicas*, 2012.
- [5] Gobierno de Navarra, “La multicausalidad del suicidio.” <http://www.navarra.es/NR/rdonlyres/4E547770-35F9-45E8-9699-169F4E0BCEE1/474575/LAMULTICAUSALIDADELSUICIDIO2.pdf>. Recuperado 23 de marzo de 2023.
- [6] Organización Panamericana de la Salud (OPS), “Marco de referencia para la implantación de la estrategia regional de salud mental.” <https://www.paho.org/hq/dmdocuments/2011/Salud-Mental-Marco-de-referencia-2011.pdf>, 2011. Recuperado 19 de abril de 2023.
- [7] Organización Panamericana de la Salud (OPS) and Organización Mundial de la Salud (OMS), “Prevención de la conducta suicida.” <https://iris.paho.org/>

REFERENCIAS

- bitstream/handle/10665.2/31167/9789275319192-spa.pdf, 2016. Recuperado 19 de abril de 2023.
- [8] B. S. Peterson, H. Zhang, R. S. Lucia, R. A. King, and M. Lewis, “Risk factors for presenting problems in child psychiatric emergencies,” *Journal of the American Academy of Child Adolescent Psychiatry*, vol. 35, no. 9, pp. 1162–1173, 1996.
- [9] J. A. Andrews and P. M. Lewinsohn, “Suicidal attempts among older adolescents: Prevalence and co-occurrence with psychiatric disorders,” *The American Academy of Child Adolescent Psychiatry*, vol. 31, no. 4, pp. 655–662, 1992.
- [10] M. Shafii, S. Carrigan, J. R. Whittinghill, and A. Derrick, “Psychological autopsy of completed suicide in children and adolescents.,” *The American journal of psychiatry*, vol. 142 9, pp. 1061–4, 1985.
- [11] Instituto Nacional de Estadística y Geografía (INEGI), “Comunicado de prensa núm. 503/22.” https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2022/EAP_SUICIDIOS22.pdf, 2022. 18, 19, 20
- [12] E. Rich and K. Knight, *Artificial Intelligence*. Artificial Intelligence Series, McGraw-Hill, 1991.
- [13] Oracle Cloud Infrastructure (OCI), “¿qué es la inteligencia artificial (ia)?.” <https://icd.who.int/browse11/1-m/es>. Recuperado 13 de marzo de 2023.
- [14] D. Poole, A. Mackworth, and R. Goebel, *Computational Intelligence: A Logical Approach*. 01 1998.
- [15] J. Muñoz Pérez, *Inteligencia computacional inspirada en la vida*. 2010.
- [16] C. M. Bishop, *Neural Network for Pattern Recognition*. Clarendon Press - Oxford, 1995.

-
- [17] S. T. y Konstantinos Koutroumbas, *Pattern Recognition*. London: Academic Press, 4 ed., 2010.
- [18] Microsoft, “Conceptos de minería de datos.” <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>, 2018. Recuperado 13 de marzo de 2023.
- [19] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.
- [20] T. M. Mitchell, *Machine learning*, vol. 1. McGraw-hill New York, 1997.
- [21] A. Geron, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. Sebastopol, CA: O’Reilly Media, 2017.
- [22] X. Zhu, “Semi-supervised learning literature survey,” Tech. Rep. 1530, Computer Sciences, University of Wisconsin-Madison, 2005.
- [23] A. Gammerman, V. Vovk, and V. Vapnik, “Learning by Transduction,” *arXiv e-prints*, p. arXiv:1301.7375, Jan. 2013.
- [24] E. M. Lara Muñoz, *Fundamentos de investigación: un enfoque por competencias*. Alfaomega, 2011.
- [25] M. Gil, “¿por qué una metodología para la gestión de proyectos?.” <https://nae.global/es/por-que-una-metodologia-para-la-gestion-de-proyectos>, 2015. Recuperado 15 de abril de 2023.
- [26] F. J. Mory, “Metodologías Ágiles vs tradicionales.” <https://revistaempresarial.com/tecnologia/metodologias-agiles-vs-tradicionales/>. Recuperado 15 de abril de 2023.
-

REFERENCIAS

- [27] W. W. Royce, “Managing the development of large software systems: concepts and techniques,” *Proc. IEEE WESTCON, Los Angeles*, pp. 1–9, August 1970.
- [28] R. S. Pressman, *Ingeniería del software: un enfoque práctico*. McGraw-Hill, 2006.
- [29] B. W. Boehm, “A spiral model of software development and enhancement,” *Computer*, vol. 21, no. 5, pp. 61–72, 1988.
- [30] J. R. Foster, *Cost factors in software maintenance*. PhD thesis, Durham University, Septiembre 1993.
- [31] S. Laoyan, “Método kaizen: la guía para la mejora continua en las empresas.” <https://asana.com/es/resources/continuous-improvement>, Octubre 2022. Recuperado 19 de abril de 2023.
- [32] J. Rodriguez, “Método kaizen: definición, pasos y ejemplos.” <https://blog.hubspot.es/sales/metodo-kaizen>, Enero 2023. Recuperado 19 de abril de 2023.
- [33] D. Anderson and A. Carmichael, *Kanban Esencial Condensado*. Blue Hole Press, 2017.
- [34] Microsoft, “Visual studio code,” 2024. Recuperado el 6 de junio de 2024.
- [35] S. Zamenfeld, “Staruml una herramienta para modelado.” <https://www.brainlabs.com/novedad/staruml-una-herramienta-para-modelado/>, 2011. Recuperado 29 de noviembre de 2023.
- [36] liderlogo.es, “¿qué es mockflow y para qué sirve en el diseño web?.” <https://www.liderlogo.es/disenio-web/que-es-mockflow/>, 2022. Recuperado 29 de noviembre de 2023.
- [37] J. Domingo Muñoz, “Qué es flask.” <https://openwebinars.net/blog/que-es-flask/>, 2017. Recuperado 30 de agosto de 2023.

- [38] J. E. Pérez, *Introducción a JavaScript*. uniwebsidad, 2008.
- [39] M. Haverbeke, *Eloquent JavaScript*. No Starch Press, 2018.
- [40] postgresql.org, “About postgresq.” <https://www.postgresql.org/about/>. Recuperado 30 de agosto de 2023.
- [41] “Servidor postgresql en azure.” <https://azure.microsoft.com/es-es/services/postgresql/>. Recuperado el 6 de junio de 2024.
- [42] “Azure app service.” <https://azure.microsoft.com/es-es/services/app-service/>. Recuperado el 6 de junio de 2024.
- [43] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” p. 1027–1035, 2007.
- [44] T. M. Kodinariya and P. R. Makwana, “Review on determining number of cluster in k-means clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90–95, 2013.
- [45] J. A. Rodrigo, “Análisis de componentes principales pca.” <https://www.cienciadedatos.net/documentos/py19-pca-python.html>, 2020. Recuperado 07 de agosto de 2023.
- [46] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, 2013.
- [47] D. W. Hosmer Jr and S. Lemeshow, “Applied logistic regression,” *Wiley Series in Probability and Statistics*, vol. 307, 1989.
- [48] D. McFadden, “Conditional logit analysis of qualitative choice behavior,” *Frontiers in Econometrics*, vol. 4, pp. 105–142, 1974.

REFERENCIAS

- [49] J. J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC, 2nd ed., 2016.
- [50] A. Agresti, *Foundations of Linear and Generalized Linear Models*. Wiley, 2015.
- [51] A. Agresti, *Categorical Data Analysis*. Wiley, 2002.
- [52] A. Agresti, *An Introduction to Categorical Data Analysis*. Wiley, 2007.
- [53] J. Smith and S. Johnson, “Understanding logistic regression: Coefficients, odds ratios, and their interpretation,” *Journal of Statistics Education*, vol. 28, no. 2, pp. 87–103, 2020.
- [54] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications Ltd, 4th ed., 2013. Chapter 17: Logistic Regression.
- [55] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 1st ed., 2006. Chapter 7: Logistic Regression Models.
- [56] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. John Wiley & Sons, 5th ed., 2012. Chapter 4: Inferences in Regression Analysis.
- [57] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer Science & Business Media, 2013.
- [58] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Regression Models*. McGraw-Hill Irwin, 2005.
- [59] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer, 2002.

- [60] G. Schwarz, *Estimating the Dimension of a Model*. Annals of Statistics, 1978.
- [61] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Applied Linear Regression Models*. McGraw-Hill Education, 2012.
- [62] J. Fox, “Applied regression analysis and generalized linear models,” *SAGE Publications*, 2015.
- [63] L. Gonick and W. Smith, *The Cartoon Guide to Statistics*. HarperPerennial, 1993.
- [64] W. Mendenhall, *Introducción A la Probabilidad Y Estadística*. Cengage Learning Latin America, 2008.
- [65] C. Ortega, “Desviación estándar: Qué es, usos y cómo obtenerla.” <https://www.questionpro.com/blog/es/desviacion-estandar/>. Recuperado 13 de febrero de 2024.
- [66] G. Westreicher, “Cuartil.” <https://economipedia.com/definiciones/cuartil.html>, 2022. Recuperado 15 de febrero de 2024.
- [67] R. García Bellido, J. González Such, and J. Jornet Meliá, *SPSS: Pruebas No Paramétricas*. 2010.
- [68] J. A. Rodrigo, “Correlación lineal y regresión lineal simple.” https://cienciadedatos.net/documentos/24_correlacion_y_regresion_lineal, 2016. Recuperado 12 de mayo de 2024.
- [69] E. Aragonès, E. Comín, M. Cavero, V. Pérez, C. Molina, and D. Palao, “Un sistema informatizado de apoyo a las decisiones clínicas para el manejo de la depresión en atención primaria,” *Atención Primaria*, vol. 49, no. 6, pp. 359–367, 2017.
- [70] C. Card, “How facebook ai helps suicide prevention.” <https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/>, 2021. Recuperado 23 de marzo de 2023.

REFERENCIAS

- [71] B. Zalar, B. Kores Plesnicar, I. Zalar, and M. Mertik, "Suicide and suicide attempt descriptors by multimethod approach," *Psychiatria Danubina*, vol. 30, pp. 317–322, 09 2018.
- [72] J. Barros, S. Morales, A. Garcia, O. Echavárri, R. Fischman, M. Szmulewicz, C. Moya, C. Núñez, and A. Tomicic, "Recognizing states of psychological vulnerability to suicidal behavior: a bayesian network of artificial intelligence applied to a clinical sample.," 08 2019.
- [73] M. Maino, S. Morales, O. Echavárri, J. Barros, A. García, C. Moya, T. Szmulewicz, R. Fischman, C. Núñez, and A. Tomicic, "Suicide risk configuration system in a clustered clinical sample: a generalized linear model obtained through the lasso technique," *Brazilian Journal of Psychiatry*, vol. 41, 10 2018.
- [74] García de la Garza, C. Blanco, M. Olfson, and M. M. Wall, "Identification of Suicide Attempt Risk Factors in a National US Survey Using Machine Learning," *JAMA Psychiatry*, vol. 78, pp. 398–406, 04 2021.
- [75] A. Bello-Valle, A. Martínez-Rebollar, W. Sánchez, and H. Estrada Esquivel, "A predictive model for automatic detection of loneliness and social isolation using machine learning," *Computación y Sistemas*, vol. 26, 03 2022.
- [76] W. Sanchez, A. Martínez-Rebollar, W. Campos, H. Estrada Esquivel, and H. Estrada, "Determinación de soledad en adultos mayores a partir de smartphones," 01 2013.
- [77] A. Tarekegn, F. Ricceri, G. Costa, E. Ferracin, and M. Giacobini, "Predictive modeling for frailty conditions in elderly people: Machine learning approaches." <https://medinform.jmir.org/2020/6/e16678>, 2020.
- [78] J. Brooke, "Sus - a quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, pp. 4–7, 1986.

- [79] J. Sauro, “Measuring usability with the system usability scale (sus),” *Measuring Usability Blog*, 2011. Recuperado el 6 de junio de 2024.
- [80] K. González, “Envejecimiento demográfico en México: análisis comparativo entre las entidades federativas,” *La situación demográfica de México 2015*, pp. 113–129, 2015.

Consentimiento informado para persona adulta mayor y responsable

Instrucciones: Favor de leer detenidamente el siguiente texto y si acepta la participación del adulto mayor que tiene a su cargo, solicitar su firma en el apartado correspondiente.

Como parte de las actividades del estudio “Modelo de predicción del aprecio corporal y la función física en adultos mayores mexicanos” que se realizará con la finalidad de detectar e implementar acciones para mejorar su estado de nutrición y calidad de vida.

1. Procedimientos: si aceptan participar en el estudio, el participante será integrado en las siguientes actividades:

a) Se realizarán entrevistas al adulto (o acompañante que le pueda auxiliar a responder) en donde se preguntará sobre el estado de salud, consumo de alimentos, aspectos psicológicos y socioeconómicos por medio de breves cuestionarios.

b) Al adulto se le realizaran mediciones de antropométricas tales como: peso, estatura, fuerza de agarre, velocidad de marcha, así como composición corporal.

2. Beneficio de participación:

a) Obtendrá información acerca del estado de nutrición del adulto mayor, conocerá si existe algún riesgo para la salud del adulto y la forma de prevenirlos.

b) La información que se obtenga permitirá implementar estrategias para mejorar el estado nutricional y psicológico en los adultos mayores.

c) El estudio no pone en ningún riesgo la salud y la vida del adulto mayor.

d) No recibirá compensación económica por su participación.

3. Posibles riesgos y molestias. Ninguno.

4. Participación voluntaria/ abandono. La participación en este estudio es de manera voluntaria, por lo que se puede retirar en cualquier momento sin consecuencia alguna. Los datos que proporcione serán secretos, lo que garantiza la confidencialidad de la persona participante.

5. Preguntas. Si tiene alguna duda, comentarios o quejas, favor de comunicarse con el Dra. Lilián Elizabeth Bosques Brugada, Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias de la Salud, Carr. Actopan –Tilcuahutla, ExHacienda la Concepción. Tel: 01-771-71-720-00, Ext. 4325, 4344, email: lilian_bosques@uaeh.edu.mx

6. Confidencialidad: Las opiniones e ideas que exprese durante la entrevista serán anónimas. Se entiende por anónimo a la condición en que el mismo investigador puede relacionar a una persona con la información.

CONSENTIMIENTO PARA PARTICIPAR EN EL PROYECTO

Los investigadores y personal del PROYECTO me han explicado y dado a conocer en qué consiste el estudio, los posibles riesgos y beneficios de la participación, entiendo que puedo dejar de participar en cualquier momento que lo desee. Me doy por enterado(a) que los resultados obtenidos en el estudio serán para beneficio de los adultos y de la institución en donde residen o participan.

Los datos que se obtengan serán sólo para fines científicos y elaborar programas de intervención para mejorar las condiciones nutricionales de los adultos mayores.

Instrumentos de medición

B.1. Cuestionario Autoestima

- Me siento una persona valiosa como las otras
- Generalmente me inclino por pensar que soy un fracaso
- Creo que tengo algunas cualidades buenas
- Soy capaz de hacer las cosas tan bien como los demás
- Creo que no tengo mucho de lo que estar orgulloso
- Tengo una actitud positiva hacia mí mismo
- En general me siento satisfecho conmigo mismo
- Me gustaría tener más respeto por mí mismo
- Realmente me siento inútil en algunas ocasiones
- A veces pienso que no sirvo para nada

B.2. Cuestionario Gratitud

- Tengo mucho en la vida por lo cual estar agradecido (a)

- Si tuviera que enumerar todo por lo que me siento agradecido(a), sería una lista muy larga
- Cuando miro al mundo, no veo mucho por lo que deba estar agradecido(a)
- Estoy agradecido con una amplia variedad de personas
- A medida que me hago mayor, me encuentro más capaz de apreciar a las personas, los eventos y las situaciones que han sido parte de la historia de mi vida
- Puede pasar mucho tiempo antes de que me sienta agradecido (a) con algo o alguien

B.3. Cuestionario Apreciación corporal

- Respeto mi cuerpo
- Me siento bien acerca de mi cuerpo
- En general me siento satisfecho(a) con mi cuerpo
- A pesar de sus defectos, acepto mi cuerpo por lo que es
- Siento que mi cuerpo tiene por lo menos algunas cualidades buenas
- Adopto una actitud positiva hacia mi cuerpo
- Soy considerado(a) con las necesidades de mi cuerpo
- Mi valor personal es independiente de mi figura o peso corporal
- No enfoco mucha energía en preocuparme por mi figura o peso corporal
- Mis sentimientos hacia mi cuerpo son positivos, en su mayor parte
- Me involucro en conductas saludables para cuidar de mi cuerpo

- No permito que imágenes poco realistas de delgadez presentadas en los medios afecten mis actitudes hacia mi cuerpo
- A pesar de sus imperfecciones, me gusta mi cuerpo

B.4. Cuestionario Ansiedad y Depresión

- Me siento tenso(a) o nervioso(a)
- Todavía disfruto con lo que antes me gustaba
- Tengo una sensación de miedo, como si algo horrible me fuera a suceder
- Puedo reírme y ver el lado divertido de las cosas
- Tengo mi mente llena de preocupaciones
- Me siento alegre
- Puedo estar sentado(a) tranquilamente y sentirme relajado(a)
- Me siento como si cada día fuera más lento
- Tengo una sensación extraña, como de “aleteo” en el estomago
- He perdido el interés por mi aspecto personal
- Me siento inquieto(a), como si no pudiera dejar de moverme
- Me siento optimista respecto al futuro
- Me asaltan sentimientos repentinos de pánico
- Me divierto con un buen libro, la radio o un programa de televisión

B.5. Cuestionario Ideación suicida

- En el último mes, ¿Qué tan seguido has tenido pensamientos suicidas?
- En el último mes, ¿Qué tan seguido has podido controlar estos pensamientos?
- En el último mes, ¿Qué tan cerca has estado de hacer un intento suicida?
- En el último mes, ¿Qué tan mal te has sentido por los pensamientos suicidas?
- En el último mes, ¿Cuánto han interferido los pensamientos suicidas con tu habilidad para llevar a cabo tus actividades de la vida diaria, como el trabajo, que haceres domésticos o actividades sociales?

B.6. Cuestionario Aprecio Funcional

- Aprecio mi cuerpo por lo que es capaz de hacer
- Estoy agradecido (a) por la salud de mi cuerpo, aunque no siempre sea tan saludable como me gustaría
- Aprecio que mi cuerpo me permita comunicarme e interactuar con los demás
- Reconozco y agradezco cuando mi cuerpo se siente bien y/o relajado
- Agradezco que mi cuerpo me permita participar en actividades que me gustan o considero importantes
- Siento que mi cuerpo hace mucho por mí
- Respeto mi cuerpo por las funciones que realiza

B.7. Cuestionario Lubben

- ¿Con cuántos familiares tiene usted contacto personal o telefónico al menos una vez por mes?
- ¿Con cuántos familiares se siente usted cómodo para conversar con facilidad sobre los asuntos privados que a usted le preocupan?
- ¿A cuántos familiares los siente usted lo suficientemente cercanos como para llamarlos en caso de necesitar ayuda?
- ¿Con cuántos amigos tiene usted contacto personal o telefónico al menos una vez por mes?
- ¿Con cuántos amigos se siente usted cómodo para conversar con facilidad sobre los asuntos privados que a usted le preocupan?
- ¿A cuántos amigos los siente usted lo suficientemente cercanos como para llamarlos en caso de necesitar ayuda?