

**Análisis de un conjunto de datos para la predicción de un ataque de corazón
utilizando clasificadores en Machine Learning.**

Yeison Alzate Quintero

Corporación Universitaria Remington.

Facultad de Sistemas.

Ingeniería en Sistemas

Rionegro Antioquia

2023



**Análisis de un conjunto de datos para la predicción de un ataque de corazón
utilizando clasificadores en Machine Learning**

Yeison Alzate Quintero

Opción de Trabajo de grado Seminario-Diplomado.

Nombre del Tutor: María Mercedes Lagos Pérez

Corporación Universitaria Remington.

Facultad de Sistemas.

Ingeniería en Sistemas

Rionegro Antioquia

2023

DEDICATORIA

Dedico este proyecto a Dios ante todo por ser siempre mi luz en todo momento, a mi madre por su inquebrantable apoyo, amor y sacrificio que han sido la base de mi camino, me ha enseñado el valor del esfuerzo y la dedicación.

Este proyecto de grado está dedicado con amor y gratitud a todos aquellos que, directa o indirectamente, han sido parte de este camino y han contribuido a mi crecimiento académico y personal.

Yeison Alzate Quintero

AGRADECIMIENTO

Agradezco a la Corporación Universitaria Remington por ayudarme a retomar mis estudios y ser parte de esta comunidad académica tan aportante en mi crecimiento personal, agradezco a todos los profesores y compañeros que han sido parte a la contribución de mi aprendizaje y conocimientos.

Un agradecimiento en especial a la docente María Mercedes Lagos Pérez por su paciencia y acompañamiento en la última etapa de mi proceso educativo, es inspiradora a realizar las cosas con amor y constancia, siempre me recordara dar el valor agregado en cualquier trabajo a realizar.

Yeison Alzate Quintero

1. TABLA DE CONTENIDO

	Pág.
2. Resumen	8
3. Palabras claves	9
4. Introducción	11
5. Planteamiento del problema	12
6. Justificación	14
7. Marco teórico	15
7.1 conjunto de datos	15
7.2 colabatory	15
7.3 python 16	
7.4 algoritmo:	16
7.5 análisis de datos.....	18
7.6 inteligencia artificial (ia).....	20
7.6.1 tipos de inteligencia artificial:	21
7.7 machine learning	21
7.8 clasificadores o aprendizaje supervisado	22
8. Objetivos	24
8.1 objetivo general	24
8.2 objetivos específicos	24
9. Metodología	25
9.1 tipo de investigación:	25
9.2 fuentes y técnicas de recolección de información	25
9.2.1 fuentes primarias.	25

9.2.2	fuentes secundarias	26
9.3	delimitación y alcance	28
9.3.1	delimitación temporal.....	28
9.3.2	delimitación espacial	28
9.3.3	alcance.....	28
9.4	población y muestra	29
9.4.1	población.....	29
9.4.2	muestra	29
10.	Resultados	30
10.1	procesar el conjunto de datos predicción ataque de corazón interpretando un algoritmo en lenguaje de programación python.	30
10.2	obtener una representación estadística a través de la interpretación de un algoritmo en python para analizar el conjunto de datos predicción ataque de corazón.....	31
10.3	interpretar un algoritmo machine learning en python para realizar la predicción de un ataque al corazón al conectar 6 inteligencias artificiales o clasificadores.....	36
11.	Conclusiones	45
12.	Referencias.....	46

TABLA DE ILUSTRACIONES.

	Pág.
Ilustración 1: Árbol del problema.....	13
Ilustración 2: Conjunto de datos.....	27
Ilustración 3: Interpretacion de algoritmo en Python que permite la conexión de Collaborative con el Conjunto de Datos.	31
Ilustración 4: Grafico de correlaciones. Implementado en Python.....	33
<i>Ilustración 5: Algoritmo en Python implementado para el grafico de correlaciones.....</i>	<i>34</i>
Ilustración 6: Tabla de frecuencia en Python para identificar el grupo de personas en riesgo de un ataque de corazón del conjunto de datos.	35
<i>Ilustración 7: Implementación de algoritmo en Python para calcular la frecuencia absoluta de la variable riesgo ataque cardiaco.....</i>	<i>36</i>
<i>Ilustración 8: Descripción de los datos para identificar los datos de entra y salida para la interpretación de un algoritmo Machin Learning en Phytton.</i>	<i>37</i>
Ilustración 9: Algoritmo implementado donde se especifica los datos de entrada y salida..	38
Ilustración 10: Identificación de las 6 inteligencias artificiales en el algoritmo de Python. Autor: Juan Carlos Briñez de León, facilitador de los algoritmos para la implementación.	41
Ilustración 11: Test de porcentaje de efectividad conectado al conjunto de datos Predicción Ataque de corazón con los 6 clasificadores. Autor. Juan Carlos Briñez de León, facilitador de los algoritmos para la implementación.	42
Ilustración 12: Código em Python para que una persona pueda ingresar sus datos, Autor, Juan Carlos Briñez de León, facilitador de los algoritmos para la implementación.	43

Ilustración 13: Algoritmo implementado para imprimí el resultado. Autor. Juan Carlos
Briñez de León.43

Ilustración 14: Prueba de ejecución de predicción ataque de corazón Autor Juan Carlos
Briñez de León, facilitador de los algoritmos para la implementación.....44

2. RESUMEN

En este proyecto de grado se implementan diversos conocimientos adquiridos durante el seminario de grado aprovechamiento de la información en contextos multidisciplinares para la inferencia de decisiones informadas, realizado entre los meses de septiembre y octubre de 2023 de manera virtual en la Corporación Universitaria Remington, sede Medellín.

Se analiza un conjunto de datos llamado Predicción Ataque Cardíaco, descargado de la página web Kaggle, se analizan sus diferentes variables para ver su relación con un ataque al corazón del cuerpo humano. Para esto se interpreta un algoritmo en lenguaje de programación Python.

Este conjunto de datos cuenta con 8763 registros y tiene como objetivo predecir la relación de las variables para determinar la probabilidad de un ataque cardíaco. (1).

Se realiza un análisis descriptivo de las variables y los datos implementando métodos estadísticos para determinar la relación que hay entre estos y un riesgo de sufrir ataque de corazón de una persona.

Al finalizar se hace conexión del conjunto de datos con un algoritmo Machine Learning de Python, conectado a 6 inteligencias artificiales o clasificadores, donde se le enseña a tomar una decisión al ingresar los datos de una persona para predecir si está en riesgo de un posible ataque al corazón.

3. PALABRAS CLAVES

Dato: Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho. (2)

Variable: Magnitud que puede tener un valor cualquiera de los comprendidos en un conjunto. (3)

Análisis: Distinción y separación de las partes de algo para conocer su composición. (4).

Ataque cardiaco: Ocurre cuando el flujo de sangre oxigenada se obstruye repentinamente en una o más de las arterias coronarias que abastecen al músculo cardíaco y una sección del músculo no puede obtener suficiente oxígeno. (5)

Implementar: Poner en funcionamiento, aplicar métodos, medidas, etc., para llevar algo a cabo. (6)

Interpretar un algoritmo: Entender cómo funcionan para posteriormente aplicar esos procedimientos en la creación de programas. (7)

Estadística: Rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades. (8)

Predicción de datos: Emplea datos históricos para predecir eventos futuros. (9)

Machine Learning: Es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (10)

Conjunto de Datos: Un conjunto de datos es una colección de datos relacionados con un tema o sector concreto. Los conjuntos de datos incluyen diferentes tipos de información, como números, texto, imágenes, vídeos y audio, y pueden almacenarse en varios formatos, como CSV, JSON o SQL. Por lo tanto, un conjunto de datos suele incluir datos estructurados con un fin específico y relacionados con un mismo tema. (11)

Python: Es un lenguaje de programación de alto nivel que se utiliza para desarrollar aplicaciones de todo tipo. A diferencia de otros lenguajes como Java o .NET, se trata de un lenguaje interpretado. (12)

Collaborative: Permite programar y ejecutar Python en un navegador web con la posibilidad de ser compartido. (13)

IA: La inteligencia artificial o IA (Artificial Intelligence, o AI en inglés) es la simulación de procesos de inteligencia humana por parte de máquinas. (14)

Clasificadores o Aprendizaje Supervisado: Es el uso de conjuntos de datos etiquetados para entrenar algoritmos que clasifiquen datos o predigan resultados de forma precisa. (15)

4. INTRODUCCIÓN

Kaggle es una página de repositorios que almacena diversos formularios, bases de datos y conjunto de datos con fines educativos y científicos. Se requiere analiza el conjunto de datos “Predicción Ataque de Corazón” descargado de la página y se deben analizaran los datos y sus diferentes variables para ver su relación con un ataque al corazón.

El conjunto de datos Predicción Ataque del Corazón utilizado para este análisis, consta de 8763 registros de pacientes de todo el mundo, culmina en una característica de clasificación binaria crucial que denota la presencia o ausencia de riesgo de ataque cardíaco, proporcionando un recurso integral para el análisis predictivo y la investigación en salud cardiovascular. (1)

Para este análisis se utilizan herramientas y un algoritmo Machin Larning estadístico en lenguaje de programación Python previamente suministrado por el docente Juan Carlos Briñez de León, de la Corporación Universitaria Remington, sede Medellín, durante la realización del seminario de grado aprovechamiento de la información en contextos multidisciplinarios para la inferencia de decisiones informadas, este algoritmo se corre desde la plataforma Collaborative conectado al conjunto de datos Predicción Ataque del Corazón con el fin que el algoritmo aprenda del conjunto de datos para determinar si una persona corre el riesgo de sufrir un ataque al corazón.

5. PLANTEAMIENTO DEL PROBLEMA

Durante la realización del seminario aprovechamiento de la información en contextos multidisciplinarios para la inferencia de decisiones informadas, que se realizara entre los meses de septiembre y octubre del año 2023 de manera virtual en la Corporación Universitaria Remington, sede Medellín, por el estudiante Yeison Alzate Quintero, estudiante de la facultad de Ingeniería en Sistemas de la Corporación Universitaria Remington, como requisito de aprobación del seminario, se solicita un análisis de los datos a un conjunto de datos llamado Predicción Ataque de Corazón, previamente descargado de la página de repositorios Kaggle, en el cual se encuentran 8763 registros de personas de diferentes partes del mundo. (1). Para realizar este análisis no se cuentan con los conocimientos necesarios.

En el análisis, interpretación y pronosticación de este conjunto de datos no hay conocimiento en la interpretación de un algoritmo en lenguaje de programación Python, lo que imposibilita la conexión y el procesamiento del conjunto de datos Predicción Ataque de Corazón.

No se tiene conocimiento en la lectura de datos lo que imposibilita la interpretación de los 8763 registros del conjunto de datos para determinar patrones de comportamientos de las variables y los datos.

Se ignoran métodos para la ejecución de datos estadísticos lo que impide la identificación y gráfica de los datos del conjunto de datos Predicción Ataque de corazón.

Falta de conocimiento en la interpretación de algoritmos Machine Learning en lenguaje de programación Python, generando incapacidad para conectar el conjunto de datos con el algoritmo de aprendizaje para realizar la predicción de un posible ataque de corazón.

Ilustración 1: Árbol del problema.



Fuente: Realizado en la aplicación Cmaptools Autoría propia.

6. JUSTIFICACION

Como requisito de aprobación del seminario de aprovechamiento de la información en contextos multidisciplinares para la inferencia de decisiones informadas, que se realiza entre los meses de septiembre y octubre del año 2023 de manera virtual en la Corporación Universitaria Remington, sede Medellín, por el estudiante Yeison Alzate Quintero, estudiante de la facultad de Ingeniería en Sistemas de la Corporación Universitaria Remington, se exige aplicar y demostrar los conocimientos adquiridos en la realización de este seminario para su culminación satisfactoriamente.

Para cumplir estos requerimientos el estudiante debe demostrar conocimientos en el análisis de un conjunto de datos, implementando las actividades y temarios de las clases, deberá procesar el conjunto de datos Predicción Ataque de Corazón interpretando algoritmos en lenguaje programación Python, obtener una representación estadística para análisis de los datos e interpretar un algoritmo Machine Learning en Python para realizar la predicción de un ataque al corazón.

Con el resultado obtenido de este proyecto se demostrará los conocimientos y habilidades adquiridas al realizar el seminario de grado como parte de su proceso académico y su futura carrera profesional.

7. MARCO TEÓRICO

7.1 Conjunto de Datos

Es una colección organizada de datos que se almacena en un sistema de almacenamiento y proporciona una estructura para buscar y acceder a la información que se desea utilizar. Es similar al contenido de una tabla en una base de datos, donde cada fila representa un registro y cada columna representa una variable. En resumen, un conjunto de datos o en inglés DataSet es la combinación de filas y columnas que conforman los datos. (16).

7.2 Colaboratory

“Colab, también conocido como "Colaboratory", es una herramienta desarrollada por Google que permite programar y ejecutar Python en tu navegador con las siguientes ventajas:

- No requiere configuración
- Acceso a GPUs sin costo adicional
- Permite compartir contenido fácilmente
- Colab puede facilitar tu trabajo, ya seas estudiante, científico de datos o investigador de IA (13).

7.3 Python

Python, creado por Guido van Rossum en 1991, es un lenguaje de programación de código abierto. Es orientado a objetos y su sintaxis se asemeja al idioma inglés, lo que facilita su lectura. Al ser interpretado, el código se convierte en bytecode y se ejecuta mediante la máquina virtual de Python. Es un lenguaje de programación versátil lo hace adecuado para diversas tareas y tiene aplicaciones en programación de instrumentos, desarrollo web, aplicaciones móviles y más.

Python es ideal para programación de servidores, operación de sistemas, software y muchos más se utiliza común mente para realizar Scripting y automatización, desarrollo de software, análisis de datos, Machine Learning e inteligencia artificial, Ciencias de blockchain. (17).

7.4 Algoritmo:

El desarrollo temprano de métodos algorítmicos se remonta a épocas ancestrales, se presume su uso en civilizaciones antiguas como la Babilonia del 2.500 a.C., y su evolución se observa en culturas posteriores como la egipcia, la griega y la india. Uno de los algoritmos más reconocidos es el de Euclides, cuyo origen se sitúa aproximadamente en el siglo III a.C.

En periodos más cercanos, en el siglo IX, el matemático persa Al-Juarismi (también conocido como al-Khwarizmi) compiló sus trabajos matemáticos en un tratado publicado

alrededor del año 825 d.C. Este aporte fue fundamental para cimentar las bases del álgebra y las matemáticas modernas, incluyendo los sistemas numéricos que empleamos hoy en día.

El compendio de Al-Juarismi abarcaba soluciones a problemas algebraicos y detallaba métodos para resolver ecuaciones lineales y cuadráticas (de primer y segundo grado respectivamente), herramientas esenciales en la programación y la resolución de problemas en ciencias, ingeniería y tecnología. Estos métodos se basaban en un conjunto de reglas y procedimientos, los cuales hoy reconocemos como algoritmos.

Respecto al término "algoritmo", se cree que proviene de la latinización del nombre de Al-Juarismi (Algorizmi), transformándose después en "algorismus". Aunque su etimología precisa no está completamente esclarecida, esta conexión histórica resalta su influencia en el desarrollo temprano de la ciencia computacional. (18).

En palabras simples son conjuntos finitos y ordenados de pasos que nos ayudan a resolver problemas o tomar decisiones. No están relacionados con los lenguajes de programación, ya que un mismo algoritmo o diagrama de flujo puede ser representado en diferentes lenguajes. En resumen, los algoritmos son una etapa previa a la programación donde se establece un ordenamiento.

Existen cuatro tipos de algoritmos en el campo de la informática: Por un lado, están los algoritmos computacionales, los cuales se resuelven mediante cálculos y pueden ser ejecutados sin dificultad por una calculadora o computadora.

Por otro lado, tenemos los algoritmos no computacionales, que no requieren de un computador para su resolución y son exclusivamente realizados por seres humanos.

Además, encontramos los algoritmos cualitativos, los cuales no involucran cálculos numéricos, sino que se basan en secuencias lógicas y/o formales para su resolución.

Por último, están los algoritmos cuantitativos, que dependen de cálculos matemáticos para llegar a su solución. (19).

7.5 Análisis de Datos

En las décadas del 30 al 49, la era de la computación se inició con hitos como la máquina de Turing y Good, que descifraba códigos enemigos durante la Segunda Guerra Mundial. Se desarrollaron máquinas como el Predictor de Kerrison para defensa antiaérea y computadoras utilizadas en el Proyecto Manhattan para prever reacciones nucleares.

Entre 1950 y 1969, la comercialización de la analítica comenzó con el primer modelo de predicción meteorológica y la resolución del problema del camino más corto. Se aplicaron modelos predictivos en riesgos de crédito y surgió el proyecto analítico del SAS Institute.

De 1970 a 1999, el análisis se popularizó con el modelo Black-Scholes para predecir precios de acciones, herramientas para sistemas de apoyo en decisiones y el surgimiento de plataformas como Amazon y eBay, impulsando la personalización en experiencias en línea.

Finalmente, en 1998, Google implementó algoritmos de búsqueda que priorizaban la relevancia de los resultados. (20)

El análisis de datos se dedica a examinar un conjunto de datos con el fin de obtener conclusiones que nos permitan tomar decisiones o ampliar nuestros conocimientos en diferentes temas. Este proceso implica realizar operaciones en los datos para obtener conclusiones precisas que nos ayuden a alcanzar nuestros objetivos. Sin embargo, estas operaciones no pueden ser definidas de antemano debido a posibles dificultades que puedan surgir durante la recolección de datos.

Si deseamos obtener datos útiles, es necesario realizar un análisis de estos. Para ello, existen diversas técnicas que se adaptan al tipo de información que se está recopilando, por lo tanto, es importante tener definida la técnica a utilizar antes de implementarla.

El análisis de datos puede ser cualitativo o cuantitativo. En el caso del análisis cualitativo, los datos se presentan de manera verbal o en ocasiones en gráficas, y se basa en la interpretación.

Por otro lado, el análisis de datos cuantitativos se basa en resultados numéricos y tangibles. Se centra en llegar a una conclusión basada únicamente en lo que ya es conocido por el investigador. Es importante asegurarse de recopilar información precisa en la que se pueda confiar, y que esté relacionada con la forma en que se planea analizar y utilizar dichos datos. Para esto, existen diversas técnicas de recolección de datos disponibles. (21).

7.6 Inteligencia Artificial (IA)

La historia de la inteligencia artificial comenzó en 1943 con Warren McCullough y Walter Pitts, quienes presentaron un modelo matemático para una red neuronal. En 1950, se creó Snarc, el primer ordenador de red neuronal, y Alan Turing publicó el Test de Turing. La conferencia de John McCarthy en 1956 marcó el nacimiento oficial de la IA.

Arthur Samuel desarrolló un software de ajedrez en 1952 y acuñó el término "Machine Learning" en 1959. Sin embargo, en los años 60 y 70, informes como el ALPAC y el informe "Lighthill" generaron dudas sobre la IA, conocido como el "primer invierno de la IA".

En 1980, el desarrollo de R1 (XCON) revitalizó la IA, seguido por un crecimiento significativo en inversiones. Los años 90 presenciaron el éxito de Deep Blue de IBM contra Gary Kasparov en ajedrez. Avances posteriores en 2008 con Google y en 2012 con Deep Learning marcaron hitos significativos.

El punto álgido fue en 2016 con AlphaGo de Google DeepMind venciendo a Lee Sedol en Go y el avance en juegos como Starcraft y Dota 2 con AlphaStar y OpenAI Five, respectivamente. En la actualidad, el Deep Learning y el Machine Learning son fundamentales en diversas aplicaciones empresariales en todos los sectores. (22).

La inteligencia artificial es la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano. Una

tecnología que todavía nos resulta lejana y misteriosa, pero que desde hace unos años está presente en nuestro día a día a todas horas.

7.6.1 *Tipos de Inteligencia Artificial:*

Los expertos en ciencias de la computación diferencian varios tipos de inteligencia artificial:

Sistemas que piensan como humanos: Automatizan tareas como la toma de decisiones, la resolución de problemas y el aprendizaje. Un ejemplo son las redes neuronales artificiales.

Sistemas que actúan como humanos: Se trata de computadoras que realizan tareas de forma similar a como lo hacen las personas. Es el caso de los robots.

Sistemas que piensan racionalmente: Intentan emular el pensamiento lógico racional de los seres humanos, se investiga cómo lograr que las máquinas puedan percibir, razonar y actuar en consecuencia. (14)

7.7 Machine Learning

Es una rama de la inteligencia artificial que permite a un sistema adquirir conocimiento a partir de los datos en lugar de depender de la programación explícita. Sin

embargo, el proceso de Machine Learning no es sencillo. A medida que el algoritmo se alimenta con datos de entrenamiento, puede generar modelos más precisos basados en esos datos. Un modelo de Machine Learning es el resultado obtenido al entrenar un algoritmo con datos. Después del entrenamiento, cuando se le proporciona una entrada al modelo, este produce una salida. Por ejemplo, si se utiliza un algoritmo predictivo, se creará un modelo predictivo. Luego, al proporcionarle datos a este modelo predictivo, se obtendrá un pronóstico basado en los datos utilizados para entrenarlo. (23).

7.8 Clasificadores o Aprendizaje Supervisado

El aprendizaje supervisado es un método fundamental en el campo del Machine Learning y la inteligencia artificial. En este enfoque, los algoritmos se entrenan utilizando conjuntos de datos etiquetados, lo que les permite clasificar datos o predecir resultados con precisión. Durante el proceso de entrenamiento, el modelo ajusta sus parámetros mediante la introducción de datos, refinando sus capacidades hasta alcanzar un nivel óptimo de ajuste, un paso esencial en la validación cruzada.

Este tipo de aprendizaje tiene un impacto significativo en la resolución de una amplia gama de problemas del mundo real. Por ejemplo, facilita la clasificación eficiente del spam en correos electrónicos, separándolo de los mensajes legítimos en la bandeja de entrada.

El funcionamiento del aprendizaje supervisado se basa en conjuntos de datos de entrenamiento que contienen información de entrada junto con los resultados deseados. Este enfoque permite que el modelo aprenda a medida que se expone a estos datos, evaluando continuamente su precisión mediante la función de pérdida y ajustándose para minimizar el error.

Dentro del aprendizaje supervisado, se pueden identificar dos tipos principales de problemas durante el proceso de minería de datos:

Clasificación: Este enfoque utiliza algoritmos para asignar datos de prueba de manera precisa a categorías específicas. Busca identificar entidades específicas en los conjuntos de datos, extrayendo conclusiones sobre cómo estas entidades deben ser etiquetadas o definidas. Algunos algoritmos comunes incluyen clasificadores lineales, máquinas de vectores de soporte (SVM), árboles de decisión, k vecinos más cercanos y bosques aleatorios.

Regresión: Aquí se analiza la relación entre variables dependientes e independientes. La regresión se emplea para realizar proyecciones, como predecir los ingresos de ventas de un negocio en función de ciertos parámetros. Algunos algoritmos populares para regresión incluyen la regresión lineal, regresión logística y regresión polinomial. (24)

8. OBJETIVOS

8.1 Objetivo General

Analizar un conjunto de datos para la predicción de un ataque de corazón utilizando clasificadores en Machine Learning.

8.2 Objetivos Específicos

- Procesar el conjunto de datos Predicción Ataque de Corazón interpretando algoritmos en lenguaje programación Python.
- Obtener una representación estadística a través de la interpretación de algoritmos en Python para analizar el conjunto de datos Predicción Ataque de corazón.
- Interpretar un algoritmo Machine Learning en Python para realizar la predicción de un ataque al corazón al conectar 6 inteligencias artificiales o clasificadores.

9. METODOLOGÍA

9.1 Tipo de investigación:

La investigación que se realizó para este trabajo es de tipo cuantitativa ya que se implementó métodos relacionados con la estadística, las matemáticas y la informática, con el objetivo de analizar las variables de un conjunto de datos, identificando factores de riesgo que puedan desencadenar un ataque al corazón. (25)

9.2 Fuentes y técnicas de recolección de información

9.2.1 *Fuentes primarias.*

Las fuentes primarias y principales para la realización de este proyecto se recolectan a través del seminario en modalidad virtual aprovechamiento de la información en contextos multidisciplinarios para la inferencia de decisiones, realizado entre los meses de septiembre y octubre de 2023 en la Corporación Universitaria Remington, sede Medellín, Antioquia, con el docente Juan Carlos Briñez de León, quien es el facilitador de las herramientas y algoritmos para la implementación.

Las clases del seminario se distribuyeron en 8 días durante 4 semanas de los meses de septiembre y octubre de 2023, en los cuales se abarcaron los siguientes temas:

- Tema: 1 Búsqueda de datos en diversas disciplinas de estudio.

- Tema: 2 Exploración computacional de datos.
- Tema: 3 Análisis correlacional de variables.
- Tema: 4 Regresiones lineales para el análisis de las variables.
- Tema: 5 Clasificadores convencionales para el aprovechamiento de variables.
- Tema: 6 Introducción a los algoritmos de IA para el análisis de variables.
- Tema: 7 Identificación de casos de estudio.
- Tema: 8 Normas técnicas para la construcción de un proyecto de grado.
- Tema: 9 Estrategias de divulgación de resultados.

Una de las labores fundamentales es analizar estos temas y datos para sacarles el máximo provecho, utilizando esa información para tomar decisiones informadas. Esto nos permite entender eventos pasados desde diversas perspectivas y así lograr resultados positivos en la toma de decisiones futuras sobre un hecho o acontecimiento a través del análisis de datos.

9.2.2 Fuentes secundarias.

La segunda fuente de información de este proyecto es el conjunto de datos Predicción Ataque de corazón, descargado del sitio web Kaggle, este conjunto de datos recopilar información de 8763 registros, este conjunto de datos resume una amplia gama de variables: Edad del paciente, Sexo: Género del paciente (Hombre/Mujer), Colesterol: Niveles de colesterol del paciente, Frecuencia cardíaca: Frecuencia cardíaca del paciente, Diabetes: si el paciente tiene diabetes (Sí/No), Historia familiar: Historia familiar de problemas relacionados con el corazón (1: Sí, 0: No), Tabaquismo: condición de fumador del paciente (1: fumador, 0: no fumador), Obesidad:

9.3 Delimitación y alcance

9.3.1 *Delimitación temporal.*

Esta investigación se realizó entre los meses de septiembre y octubre de 2023, durante la realización del seminario aprovechamiento de la información en contextos multidisciplinares para la inferencia de decisiones en la Corporación Universitaria Remington, sede Medellín, Antioquia. para analizar los datos o formulario del conjunto de datos “Predicción Ataque de Corazón”, y obtener tener un resultado con el análisis de datos de las variables y además interpretar un algoritmo en Python conectado a 6 inteligencias artificiales Machine Learning, para predecir un posible ataque al corazón.

9.3.2 *Delimitación espacial.*

La investigación se llevó a cabo dentro de la plataforma Colaborative con el conjunto de datos Predicción Ataque de Corazón.

9.3.3 *Alcance.*

Este trabajo de análisis de un conjunto de datos para la predicción de un ataque de corazón, utilizando clasificadores en Machine Learning, analiza 8763 registros de datos del repositorio Kaggle con un algoritmo Machine Learning de Python el cual posteriormente se

interpreta para que aprenda a predecir si una persona puede correr el riesgo de un ataque de corazón.

9.4 Población y muestra

9.4.1 Población.

La población con la cual se realizó este trabajo fue de un conjunto de 8763 registros de personas de todo el mundo, culmina en una característica de clasificación binaria crucial que denota la presencia o ausencia de riesgo de ataque cardíaco, proporcionando un recurso integral para el análisis predictivo y la investigación en salud cardiovascular. (1)

9.4.2 Muestra.

La muestra usada para la interpretación de este análisis son las variables y datos de un conjunto de datos que contiene 8763 registros, de la cual se sustrae la información para posteriormente realizar su análisis. (1)

10. RESULTADOS

10.1 Procesar el conjunto de datos Predicción Ataque de Corazón interpretando un algoritmo en lenguaje de programación Python.

Para procesar el conjunto de datos Predicción Ataque de corazón interpretando un algoritmo en lenguaje de programación Python, se implementan herramientas como Colab para trabajar el lenguaje de programación en Python, y algoritmos de programación. Estas herramientas son previamente suministradas por el señor Juan Carlos Briñez de León, docente durante la realización del seminario virtual aprovechamiento de la información en contextos multidisciplinarios para la inferencia de decisiones, entre los meses de septiembre y octubre del año 2023 con la Corporación Universitaria Remington, sede Medellín, Antioquia.

Los datos de análisis para la representación estadística se descargan del sitio web Kaggle, un espacio en ciencia de datos y una comunidad virtual de expertos en aprendizaje automático, apoyado por Google LLC., donde encontramos diferentes repositorios utilizados en la ciencia de datos, en el cual se aloja el conjunto de datos Predicción Ataque de Corazón, este contiene 8763 registros de personas de todo el mundo. (1)

En el conjunto de datos Predicción Ataque de Corazón para realizar el análisis de los datos, encontramos una serie de aspectos y condiciones que tienen un impacto en la salud del corazón humano, factores como la Edad, los niveles de colesterol en la sangre, el ritmo cardiaco, si una persona sufre diabetes, si se tiene historial familiar con problemas cardiacos, problemas de

obesidad, la influencia de los hábitos como fumar, consumir alcohol, actividad física, hábitos en el tipo de alimentación, problemas cardiacos previamente, el uso de medicamentos, niveles de estrés, el sedentarismo, índices de masa corporal, si sufre de triglicéridos, las horas de sueño por día, todos estos aspectos proporcionan un recurso integral para el análisis predictivo y la investigación en salud cardiovascular. (26) .

Ilustración 3: Interpretación de algoritmo en Python que permite la conexión de Colab con el Conjunto de Datos.

```
#Conectando Google Colab con el drive
from google.colab import drive
drive.mount('/content/drive')

#Importamos librerías de trabajo
import pandas as pd
import numpy as np
from google.colab import output

#Cargando datos
PAC=pd.read_csv(r'/content/drive/MyDrive/SEMINARIO/Prediccion_Ataque_de_Corazon_2.csv',sep=',')
PAC.head(10)

#Convierte datos tipo texto a números
Reemplazo_2={'Saludable':1,'Promedio':2, 'Malsano':3}
PAC['Dieta']=PAC['Dieta'].map(Reemplazo_2)
```

Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación

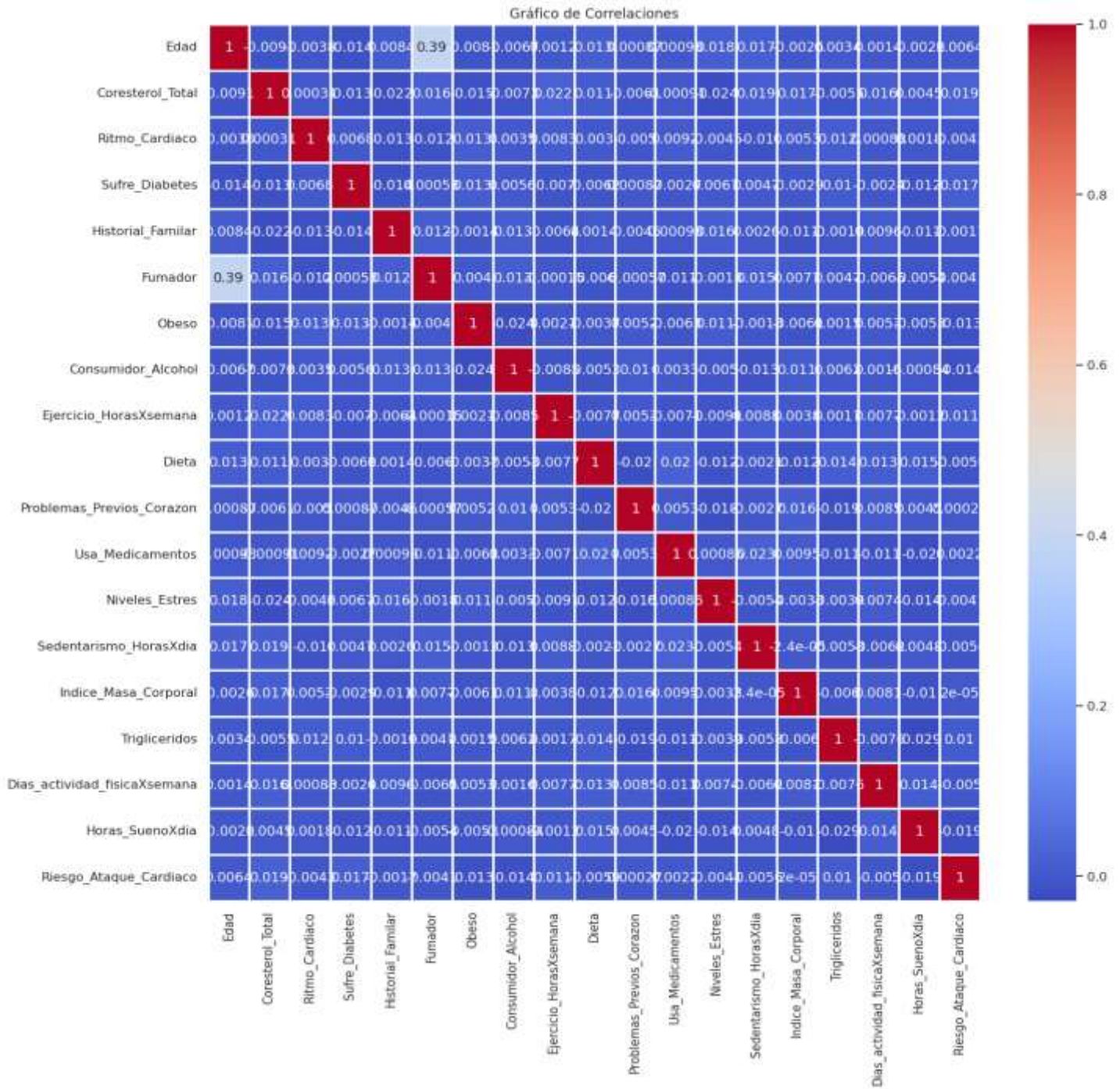
10.2 Obtener una representación estadística a través de la interpretación de un algoritmo en Python para analizar el conjunto de datos Predicción Ataque de corazón.

Con el fin de obtener una representación estadística de las variables y los datos se interpretan 2 algoritmos estadísticos en Python.

Se interpreta un algoritmo estadístico en Python llamado matriz de correlación, una matriz de correlación es una tabla que contiene los coeficientes de correlación entre diversas variables, esta matriz revela las interrelaciones entre todos los pares posibles de valores dentro de un conjunto de datos, su utilidad radica en su capacidad para condensar información y revelar patrones significativos presentes en el conjunto de datos para determinar las variables con más impacto en sufrir un ataque al corazón. (31)

El gráfico de correlaciones muestra las variables con mayor aproximación a un riesgo de ataque al corazón en una escala de 0.0 a 1.0, las variables más próximas a un valor de 1 traduce que son variables que coinciden o que tienen relación con un ataque al corazón.

Ilustración 4: Grafico de correlaciones. interpretado en Python.



Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

Al analizar los datos se encuentra que las variables que más riesgo tienen y están relacionadas con sufrir un ataque al corazón son el colesterol total 0.019, horas de sueño por día 0.019, sufrir de diabetes 0.017, consumo de alcohol 0.014, obesidad 0.013, ejercicio por semana 0.011, triglicéridos 0.01. **Ilustración 4**

Analizando en la tabla de correlaciones los valores de las variables que más relación tienen con un ataque al corazón se logró identificar como el **colesterol alto** está relacionado con: el historial familiar y el hábito del ejercicio físico; las **horas de sueño por día** con: el ejercicio físico, niveles de estrés y la dieta; **sufrir diabetes** con: la edad, el colesterol total, el historial familiar y la obesidad; **la obesidad** con: ejercicio físico, el consumo de alcohol, el colesterol total, el ritmo cardiaco, sufrir diabetes; **triglicéridos** con: la dieta, horas de sueño. **Ilustración 4**

Ilustración 5: Algoritmo en Python implementado para el grafico de correlaciones.

```
print("\n"*5)
# Matriz de correlaciones
matriz_correlaciones = PAC.corr()
sns.set(rc = {'figure.figsize':(15,15)})
# Gráfico de correlaciones
sns.heatmap(matriz_correlaciones, annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Gráfico de Correlaciones')
plt.show()
```

Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

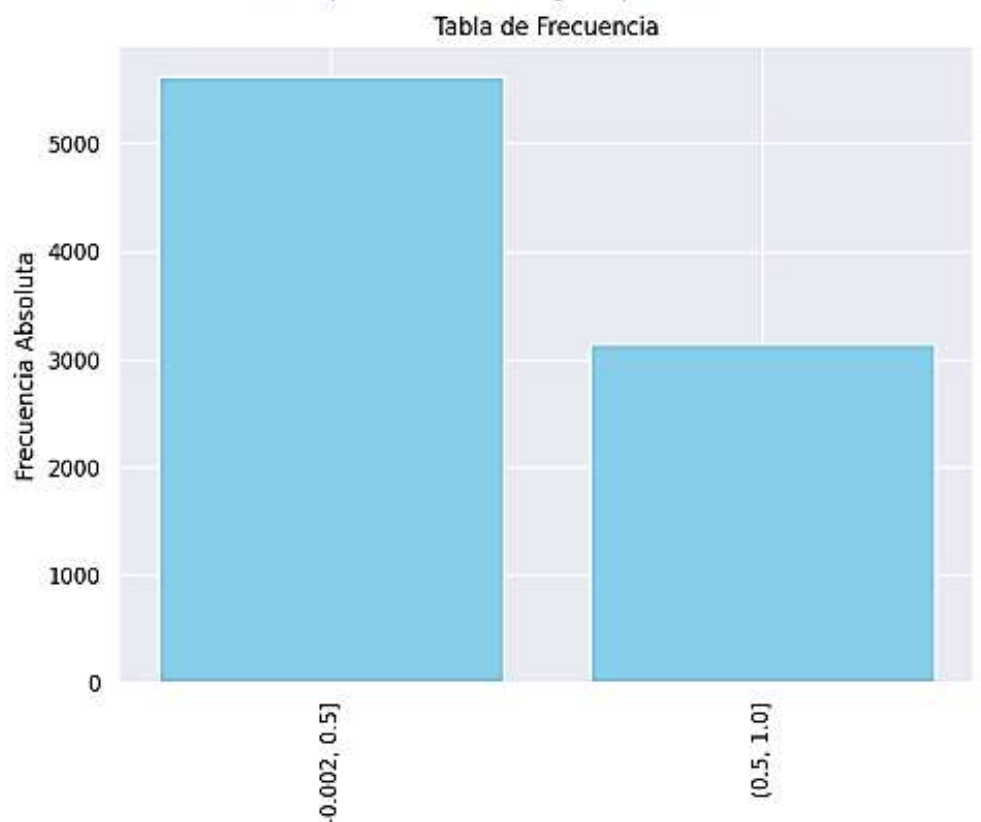
El segundo algoritmo estadístico implementado en Python es una tabla de frecuencias que organiza de manera sistemática los datos estadísticos asignándoles una frecuencia, representando la cantidad en que se repite cada número o dato en el conjunto de datos para identificar un grupo de personas de 8763 para determinar cuántas están en riesgo de sufrir un ataque al corazón.

Ilustración 6: Tabla de frecuencia en Python para identificar el grupo de personas en riesgo de un ataque de corazón del conjunto de datos.

```
Tabla de Frecuencia:
```

	Intervalo	Frecuencia Absoluta	Frecuencia Acumulada
0	(-0.002, 0.5]	5624	5624
1	(0.5, 1.0]	3139	8763

Realizando tabla de frecuencia para la variable Riesgo Ataque Cardiaco



Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

Al analizar los datos de la tabla de frecuencia absoluta que se realiza para la variable ataque cardiaco del conjunto de datos, donde el valor de 1 es alto grado de contraer problemas cardiacos y 0 pocas probabilidades, se observa que de un grupo de personas de 8763 de diferentes partes del mundo, 5624 tienen un valor entre -0.002 y 0.5 lo que significa que tienen poco riesgos de un

ataque al corazón y 3139 tienen un valor entre 0.5 y 1, estas personas tienen riesgo de sufrir un ataque al corazón. **Ilustración 6**

Ilustración 7: Interpretación de algoritmo en Python para calcular la frecuencia absoluta de la variable riesgo ataque cardiaco.

```
print("Realizando tabla de frecuencia según el número de rangos para el Riesgo de Ataque Cardíaco")
numero_de_rangos = 2 #el número de rangos es aproximadamente la raíz cuadrada del número de muestras

# Calcular la tabla de frecuencia absoluta con bins
Tabla_frecuencia = pd.cut(PAC['Riesgo_Ataque_Cardiaco'], bins=numero_de_rangos, include_lowest=True).value_counts().reset_index()
Tabla_frecuencia.columns = ['Intervalo', 'Frecuencia Absoluta']

# Calcular la frecuencia acumulada
Tabla_frecuencia['Frecuencia Acumulada'] = Tabla_frecuencia['Frecuencia Absoluta'].cumsum()

# Mostrar la tabla de frecuencia
print("Tabla de frecuencia:")
print(Tabla_frecuencia)

print("\n*9)
print("Realizando tabla de frecuencia para la variable Riesgo Ataque Cardíaco")
import matplotlib.pyplot as plt
plt.figure(figsize=(8, 6))
plt.bar(Tabla_frecuencia['Intervalo'].astype(str), Tabla_frecuencia['Frecuencia Absoluta'], color='skyblue')
plt.xlabel('Intervalo')
plt.ylabel('Frecuencia Absoluta')
plt.title('Tabla de Frecuencia')
plt.xticks(rotation=90) # Rotar etiquetas del eje x si son largas
plt.show()
```

Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

10.3 Interpretar un algoritmo Machine Learning en Python para realizar la predicción de un ataque al corazón al conectar 6 inteligencias artificiales o clasificadores.

Para realizar la predicción de un ataque al corazón se interpreta un algoritmo Machine Learning facilitado por el docente Juan Carlos Briñez de León, durante la realización del seminario aprovechamiento de la información en contextos multidisciplinarios para la inferencia de decisiones en la Corporación Universitaria Remington, Sede Medellín, Antioquia, el cual está conectado al conjunto de datos Predicción Ataque de Corazón.

Para interpretar el algoritmo Machine Learning en lenguaje de programación Python el cual se va a entrenar para tomar la decisión si una persona tiene riesgo de sufrir un ataque de corazón, primero se deben identificar los datos de entrenamiento con las 19 variables del conjunto de datos, para esto se cuenta a partir de la columna 0 a la 18, los cuales serán nuestros datos de entrada y la variable de evaluación correspondiente a Riesgo_Ataque_Cardiaco será el dato de salida la columna 18.

Ilustración 8: Descripción de los datos para identificar los datos de entrada y salida para la interpretación de un algoritmo Machine Learning en Python.

```

Describiendo datos transformados
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8763 entries, 0 to 8762
Data columns (total 19 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   Edad                                       8763 non-null   int64
1   Coresterol_Total                         8763 non-null   int64
2   Ritmo_Cardiaco                           8763 non-null   int64
3   Sufre_Diabetes                           8763 non-null   int64
4   Historial_Familiar                       8763 non-null   int64
5   Fumador                                   8763 non-null   int64
6   Obeso                                     8763 non-null   int64
7   Consumidor_Alcohol                       8763 non-null   int64
8   Ejercicio_HorasXsemana                   8763 non-null   float64
9   Dieta                                     8763 non-null   int64
10  Problemas_Previos_Corazon                 8763 non-null   int64
11  Usa_Medicamentos                          8763 non-null   int64
12  Niveles_Estres                            8763 non-null   int64
13  Sedentarismo_HorasXdia                    8763 non-null   float64
14  Indice_Masa_Corporal                      8763 non-null   float64
15  Trigliceridos                             8763 non-null   int64
16  Dias_actividad_fisicaXsemana              8763 non-null   int64
17  Horas_SuenoXdia                           8763 non-null   int64
18  Riesgo_Ataque_Cardiaco                    8763 non-null   int64
dtypes: float64(3), int64(16)
memory usage: 1.3 MB

```

Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

Con el fin de interpretar el algoritmo Machine Learning en Python indicamos los datos de extracción de la columna 0 a la 17, 18 columnas, y el dato de salida la columna 18, una columna antes de la 19 que es donde se sitúa la columna del resultado y así realizar la conexión con el conjunto de datos Predicción ataque de corazón.

Ilustración 9: Algoritmo interpretado donde se especifica los datos de entrada y salida.

```
#1. Dividing dataset into input (X) and output (Y) variables
PAC = np.array(PAC)
X = PAC[:,0:18]#Las entradas
Y = PAC[:,18]#Salida
#2. Split input and output into training and testing sets
import sklearn
from sklearn.model_selection import train_test_split
X_train, X_test,Y_train, Y_test= train_test_split(X,Y,test_size=0.001,random_state=14541)
#3. Data normalization
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
#3. Evaluando casos mediante todos los clasificadores
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score,precision_score
```

Fuente: Juan Carlos Briñez de León facilitador de los algoritmos para la interpretación.

Se utiliza 6 inteligencias artificiales o clasificadores que al conectar con las variables de entrada y salida del conjunto de datos Predicción Ataque de Corazón aprenderá a tomar una decisión si una persona tiene riesgo de sufrir un ataque al corazón.

- “**KNN**: El algoritmo de k vecinos más cercanos, también conocido como KNN o k-NN, es un clasificador de aprendizaje supervisado no paramétrico, que utiliza la

proximidad para hacer clasificaciones o predicciones sobre la agrupación de un punto de datos individual.” (27).

- **“BAYES:** El teorema de Bayes ofrece una perspectiva para evaluar y elegir entre distintos modelos al trabajar con conjuntos de datos en el ámbito del aprendizaje automático. En términos más generales, maximizar la probabilidad de que un modelo se ajuste a un conjunto de datos. Este enfoque brinda un marco probabilístico sólido para la construcción de modelos predictivos.” (28)
- **“LDA:** El Análisis Discriminante Lineal (LDA o Linear Discriminant Analysis en inglés) es una técnica de clasificación supervisada. Su objetivo radica en hallar una combinación lineal de rasgos que amplifique la distancia entre las diferentes clases presentes en un conjunto de datos de entrenamiento, puede emplearse para disminuir la dimensionalidad de un conjunto de datos, preservando la información más esencial.” (29)
- **“QDA:** Conocido como Análisis Discriminante Cuadrático, es una técnica de clasificación en el ámbito del aprendizaje supervisado. Requiere un conjunto de datos etiquetados para entrenar el modelo. Funciona estimando la distribución de probabilidad de cada clase en el espacio de características y luego emplea esta información para categorizar nuevos datos. Una característica distintiva de QDA respecto a otras técnicas, como el Análisis Discriminante Lineal (LDA), es su

capacidad para establecer límites de decisión cuadráticos entre clases, lo que lo hace más adaptable y hábil en la gestión de conjuntos de datos complejos.” (30)

- **“Tree:** Es un tipo de algoritmo de aprendizaje supervisado que se aplica principalmente en la clasificación, aunque es adaptable tanto para variables de entrada como para las de salida, ya sean categóricas o continuas. En este método, la información se divide en dos o más conjuntos homogéneos según el factor más determinante presente en las variables de entrada. El árbol de decisión identifica este factor y su valor asociado que proporciona los grupos más coherentes en la población. Se analizan todas las variables de entrada y todas las posibles divisiones, seleccionando la que ofrezca el mejor resultado.” (31)
- **“SVM:** Opera mediante la correlación de datos en un espacio de características de amplia dimensión para categorizar los puntos de datos, inclusive si no pueden separarse linealmente de otra manera. Identifica un separador entre las categorías y transforma los datos de modo que el separador se convierta en un hiperplano extraíble. Posteriormente, emplea las características de los nuevos datos para predecir el grupo al que pertenece el registro recién obtenido.” (32)

Ilustración 10: Test efectividad de las 6 inteligencias artificiales en el algoritmo de Python

```
Modelo_0 = KNeighborsClassifier(5)
Modelo_0.fit(X_train, Y_train)
Y_pred_0 =Modelo_0.predict (X_test)
print("Accuracy KNN",accuracy_score(Y_test, Y_pred_0))

Modelo_1 = GaussianNB()
Modelo_1.fit(X_train, Y_train)
Y_pred =Modelo_1.predict (X_test)
print("Accuracy Bayes",accuracy_score(Y_test, Y_pred))

Modelo_2 = LinearDiscriminantAnalysis()
Modelo_2.fit(X_train, Y_train)
Y_pred_2 =Modelo_2.predict (X_test)
print("Accuracy LDA",accuracy_score(Y_test, Y_pred_2))

Modelo_3 = QuadraticDiscriminantAnalysis()
Modelo_3.fit(X_train, Y_train)
Y_pred_3 =Modelo_3.predict (X_test)
print("Accuracy QDA",accuracy_score(Y_test, Y_pred_3))

Modelo_4 = DecisionTreeClassifier()
Modelo_4.fit(X_train, Y_train)
Y_pred_4 =Modelo_4.predict (X_test)
print("Accuracy Tree",accuracy_score(Y_test, Y_pred_4))

Modelo_5 = SVC()
Modelo_5.fit(X_train, Y_train)
Y_pred_5 =Modelo_5.predict (X_test)
print("Accuracy SVM",accuracy_score(Y_test, Y_pred_5))
```

. Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

Según la prueba de efectividad de las 6 inteligencias artificiales o clasificadores arrojé los siguientes porcentajes con cada una de ellas:

- Clasificador KNN: $0.8888 \times 100 = 88\%$
- Clasificador Bayes: $0.777 \times 100 = 77\%$
- Clasificador LDA: $0.777 \times 100 = 77\%$
- Clasificador QDA: $0.777 \times 100 = 77\%$
- Clasificador Tree: $0.555 \times 100 = 55\%$
- Clasificador SVM: $0.777 \times 100 = 77\%$

Ilustración 11: Test de porcentaje de efectividad conectado al conjunto de datos Predicción Ataque de corazón con los 6 clasificadores.

Accuracy KNN 0.8888888888888888
Accuracy Bayes 0.7777777777777778
Accuracy LDA 0.7777777777777778
Accuracy QDA 0.7777777777777778
Accuracy Tree 0.5555555555555556
Accuracy SVM 0.7777777777777778

Fuente. Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

Para determinar si una inteligencia artificial o clasificador es efectivo debe superar el 80% de efectividad, observamos que el mejor clasificador para el conjunto de datos Predicción Ataque de corazón es el KNN o también llamado vecinos más cercanos.

Por consiguiente, para ingresar los datos de una persona para que el algoritmo Machine Learning aprenda del conjunto de datos Predicción Ataque de corazón, y realice la predicción si una persona tiene poco o alto riesgo de sufrir un ataque al corazón se interpreta un algoritmo donde se declaran las nuevas variables para el registro de los datos se conecte con los 6 clasificadores para realizar la predicción.

Ilustración 12: Código interpretado en Python para que una persona pueda ingresar sus datos

```
#Probando el modelo entrenado sobre un nuevo sujeto (target)
Target=np.zeros((1,18))
Target[0,0]=float(input('Ingrese su edad: '))
Target[0,1]=float(input('Ingrese el valor de su coresterol total: '))
Target[0,2]=float(input('Ingrese su ritmo cardiaco: '))
Target[0,3]=float(input('Ingrese si sufre deabetes: 1 si y 0 no '))
Target[0,4]=float(input('Ingrese si en su familia hay antecedentes de problemas cardiacos: 1 si y 0 no '))
Target[0,5]=float(input('Ingrese si usted es fumador: 1 si y 0 no '))
Target[0,6]=float(input('Ingrese si usted es una persona obesa: 1 si y 0 no'))
Target[0,7]=float(input('Ingrese si usted consume alcohol habitualmente: 1 si y 0 no '))
Target[0,8]=float(input('Ingrese las horas de ejercicio fisico por semana '))
Target[0,9]=float(input('Ingrese su tipo de dieta: Saludable:1, Promedio:2, Malsana:3'))
Target[0,10]=float(input('Ingrese si a tenido problemas de corazon: 1 si y 0 no'))
Target[0,11]=float(input('Ingrese si usted usa medicamentos: 1 si y 0 no '))
Target[0,12]=float(input('Ingrese de 1 a 10 su nivel de estres: '))
Target[0,13]=float(input('Ingrese las horas de sedentarismo por día: '))
Target[0,14]=float(input('Ingrese el valor de indice de masa corporal: '))
Target[0,15]=float(input('Ingrese el valor de sus trigliceridos: '))
Target[0,16]=float(input('Ingrese los días de ejercicio fisico por semana: '))
Target[0,17]=float(input('Ingrese sus horas de sueño por día: '))
```

Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la interpretación.

Ilustración 13: Algoritmo interpretado para imprimir el resultado.

```
Prediction_0 = Modelo_0.predict (Target)
Prediction_1 = Modelo_1.predict (Target)
Prediction_2 = Modelo_2.predict (Target)
Prediction_3 = Modelo_3.predict (Target)
Prediction_4 = Modelo_4.predict (Target)
Prediction_5 = Modelo_5.predict (Target)
Prediction_6 = Modelo_6.predict (Target)

print(" ")

if Prediction_0<=0.0050:
    print("Según KNN, Usted tiene poco riesgo de sufrir un ataque al corazon")
else:
    print("Según KNN, Usted tiene alto riesgo de sufrir un ataque al corazon")

print(" ")

if Prediction_1==0.00005:
    print("Según Bayes, Usted tiene poco riesgo de sufrir un ataque al corazon")
else:
    print("Según Bayes, Usted tiene alto riesgo de sufrir un ataque al corazon")

print(" ")

if Prediction_2<=0.00005:
    print("Según LDA, Usted tiene poco riesgo de sufrir un ataque al corazon")
else:
    print("Según LDA, Usted tiene alto riesgo de sufrir un ataque al corazon")

print(" ")

if Prediction_3<=0.00005:
    print("Según QDA, Usted tiene poco riesgo de sufrir un ataque al corazon")
else:
```

Fuente: Juan Carlos Briñez de León.

Ilustración 14: Prueba de ejecución de predicción ataque de corazón

Ingrese su edad: 62
Ingrese el valor de su coresterol total: 310
Ingrese su ritmo cardiaco: 80
Ingrese si sufre de diabetes: 1 si y 0 no 1
Ingrese si en su familia hay antecedentes de problemas cardiacos: 1 si y 0 no 1
Ingrese si usted es fumador: 1 si y 0 no 1
Ingrese si usted es una persona obesa: 1 si y 0 no 1
Ingrese si usted consume alcohol habitualmente: 1 si y 0 no 1
Ingrese las horas de ejercicio fisico por semana 0
Ingrese su tipo de dieta: Saludable:1, Promedio:2, Malsana:32
Ingrese si a tenido problemas de corazon: 1 si y 0 no 1
Ingrese si usted usa medicamentos: 1 si y 0 no 1
Ingrese de 1 a 10 su nivel de estres: 7
Ingrese las horas de sedentarismo por dia: 13
Ingrese el valor de indice de masa corporal: 90
Ingrese el valor de sus trigliceridos: 110
Ingrese los dias de ejercicio fisico por semana: 0
Ingrese sus horas de sueño por dia: 5

Según KNN, Usted tiene alto riesgo de sufrir un ataque al corazon

Según Bayes, Usted tiene alto riesgo de sufrir un ataque al corazon

Según LDA, Usted tiene alto riesgo de sufrir un ataque al corazon

Según QDA, Usted tiene alto riesgo de sufrir un ataque al corazon

Según tree, Usted tiene alto riesgo de sufrir un ataque al corazon

Según SVM, Usted tiene poco riesgo de sufrir un ataque al corazon

Según La red neuronal, Usted tiene alto riesgo de sufrir un ataque al corazon

Fuente: Juan Carlos Briñez de León, facilitador de los algoritmos para la implementación.

11. CONCLUSIONES

En el desarrollo de este proyecto se adquirieron conocimientos necesarios para el procesamiento de un conjunto de datos y uso de herramientas como Colab que permitieron la lectura y el análisis de estos para poder implementar métodos estadísticos y la interpretación de un algoritmo Machine Learning en lenguaje de programación Python, que permitieron realizar la conexión con el conjunto de datos Predicción Ataque de Corazón para entrenar el algoritmo interpretado y realizar la predicción si una persona tiene riesgo alto o poco de sufrir un ataque de corazón.

Del conjunto de datos Predicción Ataque de Corazón se pudo llegar a la conclusión a través de una tabla de correlaciones que las variables que más relación tienen con sufrir un ataque al corazón son colesterol total, horas de sueño por día, sufrir de diabetes, consumo de alcohol, obesidad, ejercicio por semana, triglicéridos. **Ilustración 4**

Todas estas variables están correlacionadas entre sí y cada una tiene influencia la una sobre la otra en mayor o menor medida, podemos observar cómo mantener unos hábitos de vida sanos, una dieta balanceada, ejercicio regularmente y dormir lo suficiente, mejoran considerablemente la salud cardíaca. **Ilustración 4**

Según el gráfico de frecuencias que se usó para analizar la variable Riesgo Ataque Cardíaco del conjunto de datos Predicción Ataque de corazón se observa que de las 8763 personas que participaron, 3139 están en riesgo de sufrir un ataque al corazón. **Ilustración 6**

La inteligencia artificial o clasificador que más porcentaje de efectividad que arrojaron las pruebas es el clasificador KNN o también llamado vecinos más cercanos, concluimos que este clasificador es el más efectivo en la implementación de la predicción Ataque de corazón.

Ilustración 11.

Finalmente se analizó un conjunto de datos para la predicción de un ataque de corazón utilizando clasificadores en Machine Learning.

12. REFERENCIAS

BIO. <https://www.kaggle.com>. [Online].; 2023 [cited 2023 octubre 07 [El sitio web . Kaggle es un espacio científico en ciencia de datos, cuentan con una variedad de repositorios incluyendo el conjunto de datos Predicción Ataque de Corazón usado para realizar este análisis.]. Available from:

<https://www.kaggle.com/datasets/iamsouravbanerjee/heart-attack-prediction-dataset/data>.

Española RA. Real Academia Española. [Online].; 2023 [cited 2023 octubre 13 . [Definición de Dato según la Real Academia Española]. Available from:

<https://dle.rae.es/dato>.

Española RA. Real Academia Española. [Online].; 2023 [cited 2023 octubre 11 . [Definición de variable según la Real Academia Española]. Available from:

<https://dle.rae.es/variable>.

Real Academia Española. Real Academia Española. [Online].; 2023 [cited 2023 . octubre 10 [Definición de análisis según la Real Academia Española]. Available from:

<https://dle.rae.es/an%C3%A1lisis>.

Biblioteca Nacional de Medicina. Medlineplus. [Online].; 2023 [cited 2023 octubre . 10 [Definición ataque cardiaco según el sitio web biblioteca nacional de medicina

MedlinePlus]. Available from: <https://medlineplus.gov/spanish/ency/article/000195.htm>.

Real Academia Española. Real Academia Española. [Online].; 2023 [cited 2023 . octubre 13 [Definición de Implementar según la Real Academia Española]. Available

from: <https://www.rae.es/drae2001/implementar>.

Banco de Recursos de Andalucía. Banco de Recursos de Andalucía. [Online].; 2023 . [cited 2023 octubre 10 [Este sitio web es un banco de recursos educativos donde se observa

la definición de Interpretación de un Algoritmo]. Available from:
https://edea.juntadeandalucia.es/bancorecursos/file/0ae7c396-1ae9-4b65-9873-1eeadfd54623/1/es-an_2019071712_9155629.zip/descripcin_de_la_tarea.html?temp.hn=true&temp.hb=true#:~:text=Interpretar%20un%20algoritmo%20implica%2C%20al,para%20diferentes%20valore.

Real Academia Española. Real Academia Española. [Online].; 2023 [cited 2023 octubre 10 [Definición de estadística según la Real Academia Española]. Available from:
<https://dle.rae.es/estad%C3%ADstico>.

Work M. Math Work. [Online].; 2023 [cited 2023 octubre 10 [Definición de predicción de datos según el sitio web Math Works]. Available from:
<https://es.mathworks.com/discovery/predictive-analytics.html>.

IBM. IBM. [Online].; 2023 [cited 2023 octubre 10 [Una descripción más detallada sobre los algoritmos Machine Learning de la empresa tecnológica IBM]. Available from:
<https://www.ibm.com/mx-es/topics/machine-learning>.

Data B. Brigh Data. [Online].; 2023 [cited 2023 octubre 10 [Una ampliación un poco más detallada al término conjunto de datos en relación a la informática según el sitio web tecnológico Brigh Data]. Available from: <https://brightdata.es/blog/datos-web/what-is-a-dataset>.

Santander Open Academy. [Online].; 2023 [cited 2023 octubre 10 [Definición del lenguaje de programación Python según la universidad Santander Open Academy]. Available from: <https://www.santanderopenacademy.com/es/blog/python-que-es.html>.

Google. Collaborative. [Online].; 2023 [cited 2023 octubre 10 [Collaborative es un
3. espacio donde se puede escribir y ejecutar Python, creado por google]. Available from:
<https://colab.research.google.com/?hl=es>.

Iberdrola. Iberdrola. [Online].; 2023 [cited 2023 octubre 09 [Una definicion y los
4. tiposde linteligencia artificial según el sitio web de innovación en energias renovables
Iberdrola]. Available from: <https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial>.

IBM. IBM. [Online].; 2023 [cited 2023 octubre 13 [Definición de clasificadores o
5. apredizaje supervisado segun la empresa IBM]. Available from: <https://www.ibm.com/mx-es/topics/supervised-learning>.

Solis DC. Open Webinars. [Online].; 2023 [cited 2023 octubre 09 [Definicion de
6. conjunto de datos segun la web Open Webinars dedicados a la formación tecnológica.].
Available from: <https://openwebinars.net/blog/datasets-que-son-y-como-acceder-a-ellos/#:~:text=Qu%C3%A9%20es%20un%20Dataset&text=Es%20b%C3%A1sicamente%20el%20contenido%20de,posibles%20variables%20que%20la%20conforman>.

Hubspoot. Hubspoot. [Online].; 2023 [cited 2023 octubre 10 [Descripcion de Python
7. como lenguaje de prgramacion segun el sitio web Hubspoot quienes son desarrolladores y
comercializadores de productos de software]. Available from:
<https://blog.hubspot.es/website/que-es-python>.

Telefonica Tech. [Online].; 2023 [cited 2023 Octubre 11 [Esta es una breve reseña
8. historica de los algoritmos y su evolucion a traves de calculos matematicos]. Available from:

<https://telefonicatech.com/blog/breve-historia-algoritmos-matematico-siglo-ix-que-sento-bases-inteligencia-artificial>.

Concepto. Concepto. [Online].; 2023 [cited 2023 octubre 09 [Definicion de Algoritmo
9. segun enciclopedia online Concepto]. Available from: <https://concepto.de/algoritmo-en-informatica/#ixzz8Fgh5eesD>.

Data B. Bg Data. [Online].; 2023 [cited 2023 octubre 11 [Breve reseña historica del
0. analisis de datos segun el sitio web Big Data]. Available from:
<https://bigdataparacuriosos.wordpress.com/historia-big-data/>.

QuestionPro. QuestionPro. [Online].; 2023 [cited 2023 octubre 09 [Una descripcion
1. del concepto Analisis de Datos según el sitio web QuestionPro quienes se dedican a la
faricación de encuestas]. Available from: <https://www.questionpro.com/es/analisis-de-datos.html>.

Datascientest.. Datascientest. [Online].; 2023 [cited 2023 octubre 12. Available
2. from: <https://datascientest.com/es/inteligencia-artificial-definicion>.

IBM. IBM. [Online].; 2023 [cited 2023 octubre 09 [Una definicon concreta de
3. Machine Learning segun la empresa tecnologica IBM]. Available from:
<https://www.ibm.com/mx-es/analytics/machine-learning>.

IBM. IBM. [Online].; 2023 [cited 2023 octubre 13 [Explicacion de los conceptos
4. clasificadores y aprendizaje supervisado segun la empresa informatica IBM]. Available
from: <https://www.ibm.com/es-es/topics/supervised-learning>.

Masters Ty. Tesis y Masters. [Online].; 2023 [cited 2023 octubre 10. Available from:
5. <https://tesisymasters.com.co/tipos-de-investigacion/>.

Institute TTH. The Texas Heart Institute. [Online].; 2023 [cited 2023 octubre 13

6. [Segun The Texas Heart Institute menciona estos principales factores de riesgo en la salud cardiovascular]. Available from: <https://www.texasheart.org/heart-health/heart-information-center/topics/factores-de-riesgo-cardiovascular/>.

IBM. IBM. [Online].; 2023 [cited 2023 noviembre 26 [Deficnion del algoritmo de

7. machine learning KNN una inteligencia artificial de aprendizaje]. Available from: <https://www.ibm.com/mx-es/topics/knn>.

BigData.com. BigData.com. [Online].; 2019 [cited 2023 octubre 15 [Es importante

8. mencionar que para el funcionamiento de esta inteligencia artificial es necesario la coneccion a un conjunto de datos dado para el aprendizaje automático]. Available from: <https://sitiobigdata.com/2019/12/24/teorema-de-bayes-en-machine-learning/#:~:text=El%20teorema%20de%20Bayes%20proporciona,la%20intuici%C3%B3n%20a%20menudo%20falla.>

InteractiveChaos. InteractiveChaos. [Online].; 2023 [cited 2023 octubre 13

9. [Descripcion de la inteligencia artificial LDA segun el sitio web InteractiveChaos]. Available from: <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/linear-discriminant-analysis-lda>.

Ts2. Ts2. [Online].; 2023 [cited 2023 octubre 13 [Explicacion de la inteligencia

0. artificial QDA segun el sitio web Ts2]. Available from: <https://ts2.space/es/ai-y-analisis-discriminante-cuadratico-qda/#gsc.tab=0>.

AprendeIA. AprendeIA. [Online].; 2023 [cited 2023 octubre 13 [Explicación de la
1. inteligencia artificial Tree según el sitio web AprendeIA]. Available from:
<https://aprendeia.com/aprendizaje-supervisado-decision-tree-classification/>.

IBM. IBM. [Online].; 2023 [cited 2023 octubre 13 [Descripción de la inteligencia
2. artificial SVM según la empresa IBM]. Available from: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>.

Google. Colaboratory. [Online].; 2023. Available from:
3. https://colab.research.google.com/?hl=es#scrollTo=5fCEDCU_qrC0.

Corazon FE. Fundación Española del Corazon. [Online].; 2023. Available from:
4. <https://fundaciondelcorazon.com/prevencion/riesgo-cardiovascular/colesterol.html>.

Organización Panamericana de la Salud. Organización Panamericana de la Salud.
5. [Online].; 2021 [cited 2023 octubre 10 [Según la organización panamericana de la salud].
Available from: <https://www.paho.org/es/noticias/29-9-2021-enfermedades-corazon-siguen-siendo-principal-causa-muerte-america#:~:text=A%20nivel%20mundial%2C%20cada%20a%C3%B1o,donde%20los%20casos%20siguen%20aumentando>.

Salud OMD. Organización Mundial de la Salud. [Online].; 2023 [cited 2023 octubre
6. 08 [Es importante observar como las enfermedades cardio vascular es una de las más grandes
problemáticas de salud a nivel mundial]. Available from: https://www.who.int/es/health-topics/cardiovascular-diseases#tab=tab_1.

Heart N. National Heart. [Online].; 2023 [Información más detallada sobre un ataque
7. al corazón según Heart, National.]. Available from:
<https://www.nhlbi.nih.gov/es/salud/ataque-cardiaco>.

Institute TTH. The Texas Heart Institute. [Online].; 2023 [cited 2023 octubre 09
8. [Información de las causas y lo que sucede en un ataque al corazón según The Texas Heart
Institute]. Available from: <https://www.texasheart.org/heart-health/heart-information-center/topics/ataque-cardiaco/>.

QuestionPro. QuestionPro. [Online].; 2023 [cited 2023 octubre 13 [en el sitio web
9. QuestionPro encontramos información más detallada sobre la matriz de correlación].
Available from: <https://www.questionpro.com/blog/es/matriz-de-correlacion/>.