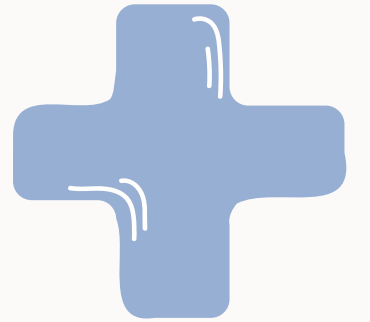


PREDICCIÓN DE ENFERMEDADES CARDÍACAS



CENTERS FOR DISEASE

ÍNDICE

1

Introducción

2

Objetivos

3

**Orígenes de
datos**

4

**Principales
indicadores**

5

**Algoritmos de
clasificación**

6

Tareas realizadas

7

Conclusiones

8

Integrantes

INTRODUCCIÓN



INTRODUCCIÓN

Según estudios realizados por los Centros para el Control y Prevención de Enfermedades (CDC) una de las principales causas de muerte en las personas en los EE.UU es debido a enfermedades cardíacas. La detección y prevención de factores de riesgo en una etapa temprana pueden salvar muchas vidas. La aplicación de métodos de aprendizaje automático para detectar patrones en base a los datos pueden incidir positivamente en la salud de la población.



CENTERS FOR DISEASE

OBJETIVOS



OBJETIVOS

El objetivo de esta investigación es poder predecir la probabilidad de que una determinada persona pueda contraer una enfermedad cardiaca, dependiendo de distintos factores tales como: la edad, el promedio de horas de descanso, la obesidad (IMC), si es fumador, si bebe alcohol, entre otras.



ORÍGENES DE DATOS



ORÍGENES DE DATOS

Encuesta anual 2020

El conjunto de datos proviene de los CDC y es una parte importante del Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS), que realiza encuestas telefónicas anuales para recopilar datos sobre el estado de salud de los residentes de EE. UU. Fue establecido en 1984 con 15 estados, BRFSS ahora recopila datos en los 50 estados. El conjunto de datos más reciente (al 15 de febrero de 2022) incluye datos de 2020. Los datos con los que trabajaremos provienen de la encuesta anual de los CDC de 2020 de 400 000 adultos relacionados con su estado de salud.



PRINCIPALES INDICADORES



PRINCIPALES INDICADORES

Índice masa corporal (IMC).

Horas de descanso

Fumador

Alcoholismo

Actividad física

Antecedentes de
enfermedades



ALGORIMOS DE CLASIFICACIÓN



Árbol de decisión

El algoritmo de árboles de decisión es un algoritmo de clasificación para su uso en el modelado predictivo de atributos discretos y continuos. Para los atributos discretos, el algoritmo hace predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos.

Random Forest

En Random Forest se ejecutan varios algoritmos de árbol de decisiones en lugar de uno solo. Para clasificar un nuevo objeto basado en atributos, cada árbol de decisión da una clasificación y finalmente la decisión con mayor "votos" es la predicción del algoritmo.

KNN (Vecinos próximos)

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.

TAREAS REALIZADAS



TAREAS REALIZADAS

1

CARGA DEL SET DE DATASET

A lo largo de este análisis se realizó la carga del set de datos original. Se creó un DF el cual se analizó sus variables, las métricas que tenía. Se realizó un trabajo de limpieza y transformación sobre los datos.

2

APLICACIÓN DE ALGORITMOS DE MACHINE LEARNING

Luego se aplicaron los algoritmos de machine learning los cuales fueron:

1. Árbol de decisión
2. Bosque aleatorio
3. Vecinos cercanos

3

PREDICCIONES

Una vez realizado el entrenamiento de los modelos con el 70% de los datos, se realizó una predicción con el 30% restante.

4

CONCLUSIONES

A partir de los resultados obtenidos se compararon los resultados.

CONCLUSIONES



CONCLUSIONES

El algoritmo que mejor resultados obtuvo fue el Random Forest, que arrojó los siguientes resultados:

- 57136 casos donde se clasifican como no cardíaco y son no cardíaco.
- 4965 casos se clasifican como No cardíaco y son cardíaco.
- 1201 casos se clasifican como cardíaco, y no son cardíaco.
- 657 casos se clasifican como cardíaco y son cardíaco.



INTEGRANTES

QUE REALIZARON ESTA INVESTIGACIÓN



EQUIPO ENCARGADO DE LA INVESTIGACIÓN



NICOLÁS HERRERA

Data Engineer



NICOLÁS BALBIANI

Data Sciences



MATÍAS VITOLA

Data Analytics

MUCHAS GRACIAS

