



## INTRODUCCIÓN

Las enfermedades cardiovasculares son la principal causa de defunción en el mundo y, según estimaciones, se cobran 17.9 millones de vida cada año.

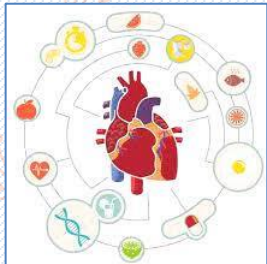


Figura 1

Las personas con enfermedad cardiovascular o que se encuentran en alto riesgo cardiovascular (debido a la presencia de uno o más factores de riesgo como hipertensión, diabetes, hiperlipidemia o enfermedad ya establecida) necesitan una detección y manejo precoces donde un modelo de aprendizaje automático puede ser de gran ayuda.

## OBJETIVO

Pronosticar que tan cerca está una persona de padecer una enfermedad cardiovascular dependiendo de si padece o no algunas de las enfermedades que se muestran dentro de la base.

### Objetivo Secundario

- Si soy fumador, ¿qué probabilidad hay de que padezca una enfermedad cardiovascular?
- Al padecer diabetes, ¿mueren más hombres que mujeres por alguna enfermedad cardiovascular?

## RECURSOS



Figura 2



Figura 3

Librerías:



## REFERENCIAS

Base de Datos:



Notebook:



## CONCLUSIONES

Gracias al apoyo de la técnica del árbol de decisión, pudimos realizar una mejor observación a los datos y cómo estos se relacionan entre sí. De acuerdo con nuestras preguntas de interés, podemos concluir que encontramos que ser fumador y padecer anemia son los factores que más impacto tienen a la salud de los pacientes dentro de esta base. Además, encontramos estos dos últimos factores de gran importancia, más aparte, que dentro de nuestra base se cuenta con una mayor cantidad de pacientes que murieron, a los que aún siguen con vida.

Algunas ideas para mejorar aun más nuestro trabajo a futuro son: tratar de implementar alguna otra técnica que pueda mostrar más claramente la supervivencia, darle un mejor formato a gráficos e información para facilitar su lectura y por último, búsqueda y realización de nuevos objetivos secundarios.

## METODOLOGÍA

La base de datos "Heart Failure Prediction" es del año 2020, cuenta con **299 registros** de pacientes donde 194 son hombres y 105 mujeres, y 13 variables para trabajar, dichas variables son:

- ☐ **Anemia:** Disminución de glóbulos rojos.
- ☐ **Diabetes:** Si el paciente tiene Diabetes.
- ☐ **Hipertensión:** Si el paciente tiene Hipertensión.
- ☐ **Plaquetas:** Plaquetas en la sangre (kiloplaquetas/mL)
- ☐ **Sexo:** Género de la persona (1 hombre y 0 mujer)
- ☐ **Fumador:** Si el paciente es fumador activo.
- ☐ **Muerte:** Si el paciente falleció en el periodo de observación.

En la limpieza de nuestros datos decidimos mantener la base de datos completa ya que no contábamos con datos nulos y toda la información es relevante para nuestro análisis.

Además, como resultado de las estadísticas básicas obtuvimos que la **edad promedio** es de **60 años** y el numero de pacientes que murieron durante el periodo de observación fue de 96.

Se analizaron los datos mediante **Árbol de Decisión**, la elección de esta técnica se debió a que dentro de nuestra base se trabaja con números booleanos de 0 y 1 que responden a si los pacientes padecen de algunas enfermedades, su sexo y si estos han fallecido, por lo que encontramos esta técnica adecuada para nuestro análisis.

Hicimos uso de la probabilidad condicional para el manejo de el árbol donde primero se calculó la probabilidad de morir teniendo o no cierta enfermedad de las mencionadas anteriormente y dado esto se realizó una codificación de la probabilidad condicional para lograr obtener la probabilidad de que alguien muera dependiendo de si tiene o no una o más enfermedades.

## RESULTADOS

Gracias al árbol de decisión y la probabilidad condicional nos es posible saber la probabilidad de muerte dependiendo de las enfermedades que se padezcan o no, como por ejemplo: la probabilidad de morir teniendo sólo diabetes en hombres es de 57.2916% y en mujeres es de 51.458%

LOS PACIENTES Y SUS ENFERMEDADES

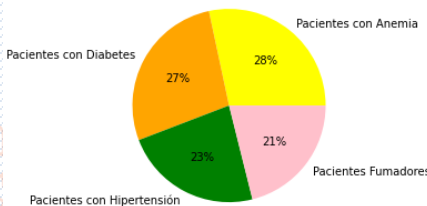


Figura 4

Figura 4. Podemos observar el porcentaje de pacientes con cada una de las enfermedades mostradas en la información, siendo así 129 con Anemia, 125 con Diabetes, 105 con Hipertensión y 96 Fumadores

Figura 5. Al comenzar a trabajar con nuestra técnica de minería, separamos las variables por independientes (Anemia, Diabetes, Hipertensión, Sexo y Fumador) y las dependientes.

Después de la obtención de nuestro Árbol de decisión, sabemos que si los pacientes murieron padeciendo hipertensión, entonces es probable que hayan sido fumadores. Si, por el contrario, no la padecieron, pasaríamos con otro factor probable. Si los pacientes que no padecieron de hipertensión padecen de diabetes, entonces pasamos a separar la muestra por género.

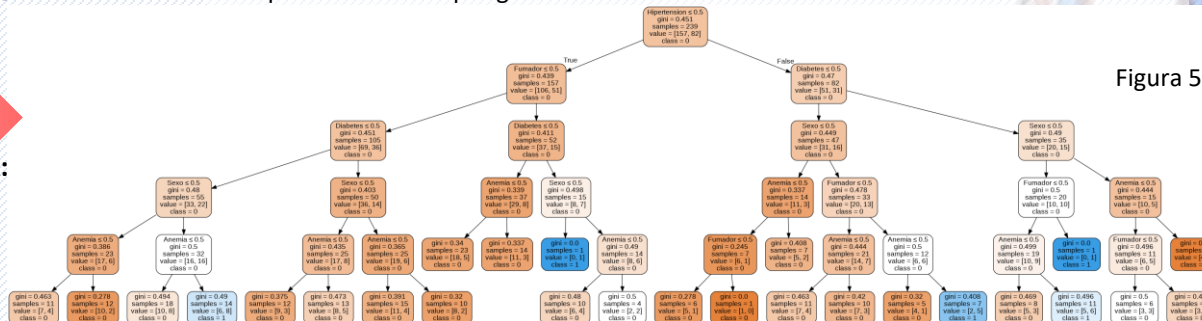


Figura 5