

PRESENTACIÓN DE PROYECTO FINAL DATA SCIENCE ¿PODREMOS OPTIMIZAR EL SISTEMA DE SALUD?

POR YORI MICAELA

ÍNDICE

- Contexto y audiencia
- Hipótesis
- Metadata
- Análisis exploratorio
- Insights y Recomendaciones
- Implementación de modelos de aprendizaje
- Conclusión

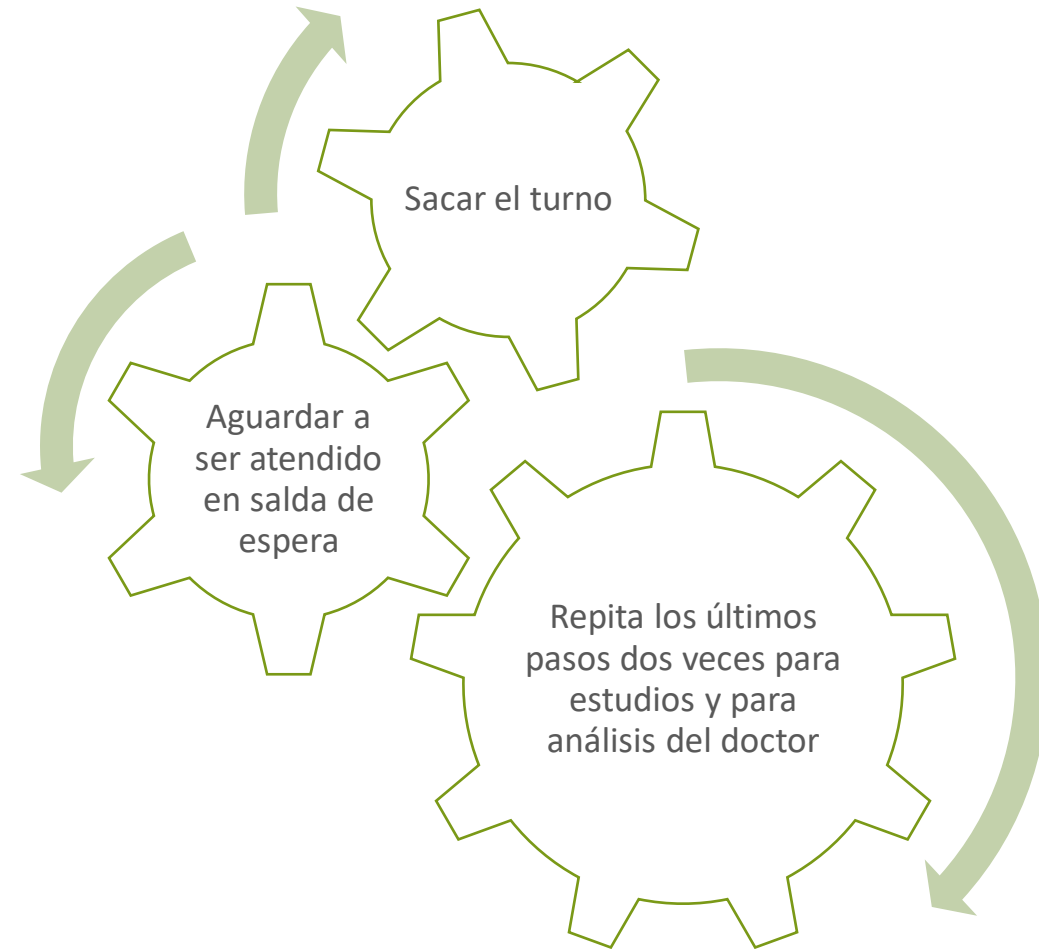
CONTEXTO

PROBLEMA

Todo el que haya asistido al médico en Argentina, tanto sea con una prepaga, una obra social o en el ámbito estatal ha tenido la experiencia de solicitar un turno con un especialista con un mes, dos, seis, hasta un año de demora.

Luego de la espera al momento del turno, el mismo no se cumple el horario pactado debido a los sobre turnos y retrasos.

En la consulta generalmente derivan al paciente a estudios y luego a volver con el medico a que los analice en una segunda visita, por lo que se repite el circulo nuevamente.



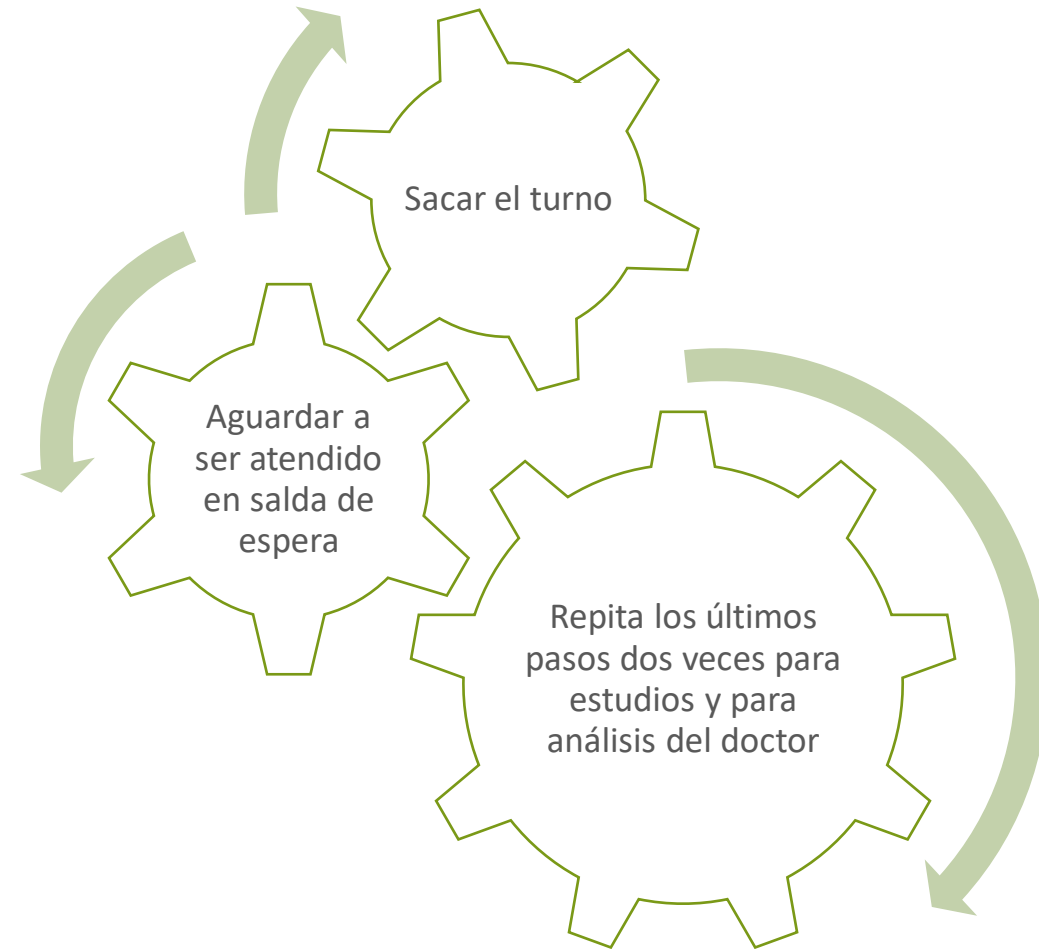
PROBLEMA

En el ámbito médico el tiempo es tirano por lo que optimizar los tiempos de atención es primordial para tratar el diagnostico de manera temprana y que el paciente no se olvide o pierda el interés.

En el presente proyecto buscamos clasificar a un individuo como enfermo o sano a partir de las variables estudiadas para así eliminar o acotar la demanda de turnos médicos.

La base de datos contiene información sobre una muestra de individuos al azar y la presentación o no de enfermedades cardíacas.

De ser exitoso el proyecto podremos diagnosticar al paciente sin necesidad de la segunda visita al cardiólogo. Lo cual en individuos sanos finalizaría la atención. En el diagnostico “Sano” nos interesa, entonces, lograr la mayor efectividad posible.



HIPÓTESIS

HIPÓTESIS

- Valores altos de presión arterial sistólica y la presión arterial diastólica condicionan enfermedades cardíacas.
- Presencia de colesterol en relación al índice de masa corporal, variables controlables con la dieta, son factores influyentes en el diagnostico.
- El alcoholismo y el tabaquismo son factores influyentes en el diagnostico de enfermedad.
- La actividad física según la edad implica menor incidencia de enfermedad.

METADATA

METADATA

- Edad medida en años
- Altura en centímetros
- Peso en kilogramos
- Género en hombre o mujer
- Presión sistólica en Milímetros de Mercurio
- Presión diastólica en Milímetros de Mercurio

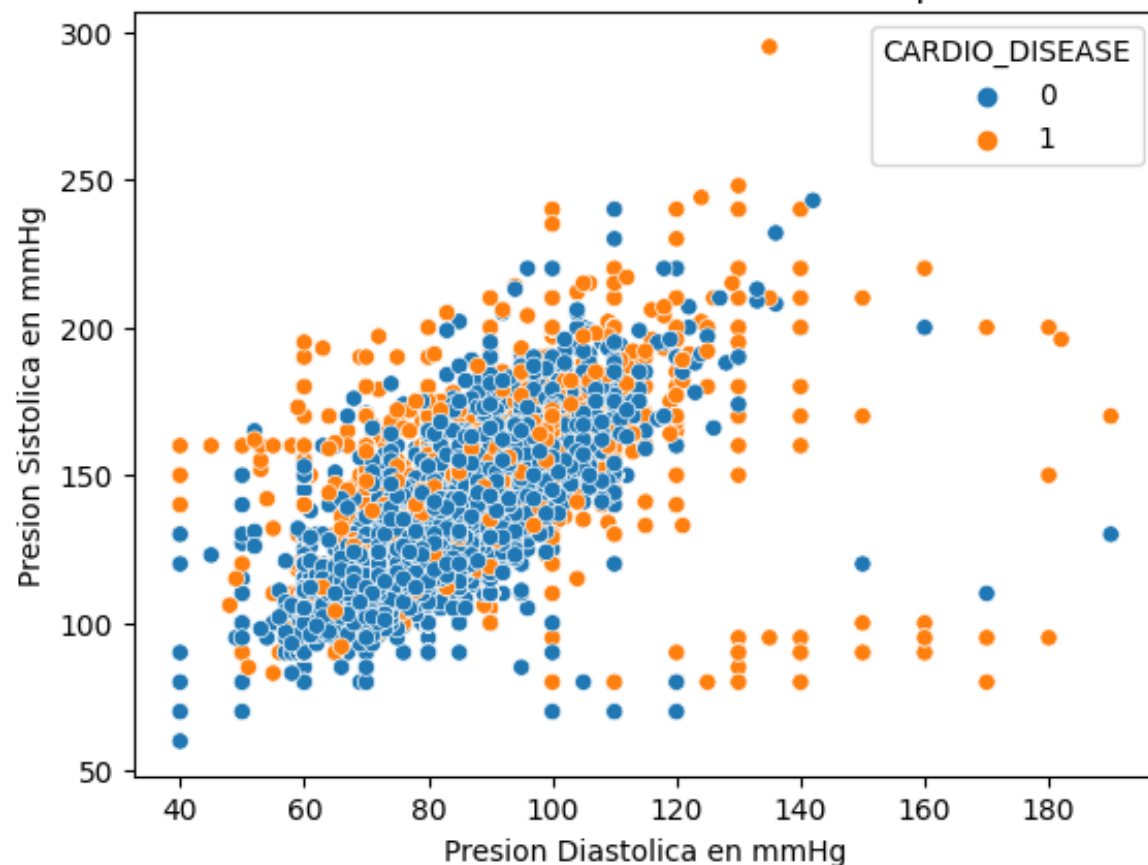
DISPONEMOS DE INFORMACIÓN SOBRE 68783 PACIENTES

- Colesterol categorizado en normal, poco normal y muy poco normal
- Glucosemia categorizado en normal, poco normal y muy poco normal
- Habito de fumador en sí o no
- Alcoholismo en sí o no
- Físicamente activo en sí o no

ANÁLISIS EXPLORATORIO

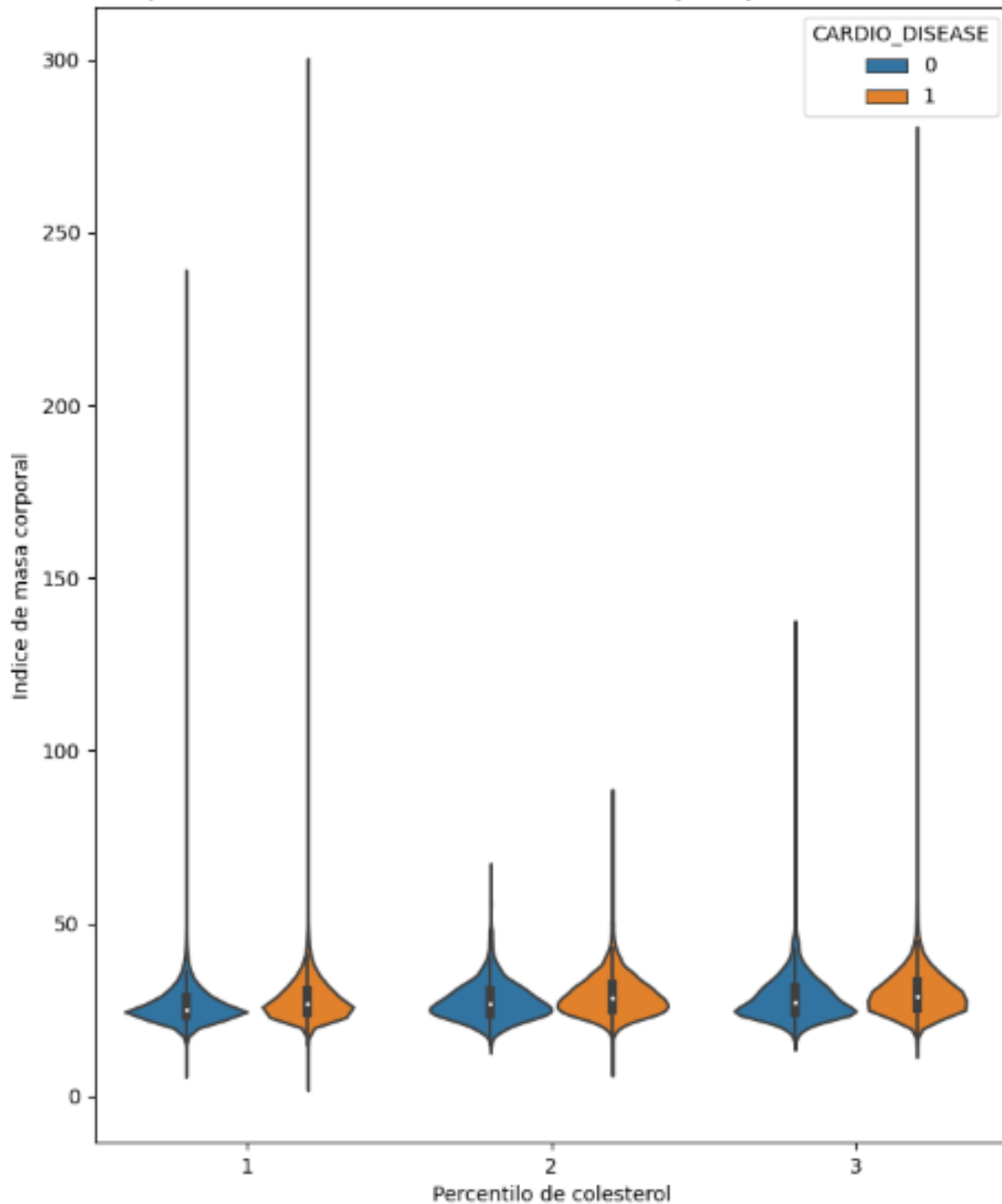
VALORES ALTOS DE PRESIÓN ARTERIAL SISTÓLICA Y LA PRESIÓN ARTERIAL DIASTÓLICA CONDICIONAN ENFERMEDADES CARDÍACAS.

Relación Presion Diastolica/Presion Sistolica discriminado por enfermo o no enfermo



- A partir del gráfico anterior podemos inducir que los individuos de presión sistólica (eje y) por encima de 125 tienen una mayor tendencia a expresar una enfermedad cardíaca siendo que 120 es el valor "normal".
- En cuando a presiones diastólicas (eje x), la mayor concentración de enfermos se haya entre 80 y 100, pese a que también son los valores considerados como "normal" lo cual nos indica que por sí solo no es un factor determinante.
- En situaciones donde tanto la presión sistólica como la diastólica tienen valores por encima de lo normal baja la probabilidad de ser un individuo sano.
- En individuos con ambas variables por debajo de lo normal no presentan enfermedades cardíacas en su mayoría.
- En casos donde la presión diastólica esta por debajo de 80-100 y la sistólica por encima 130, todos los individuos se encontraban enfermos.
- Y por ultimo, los individuos que presentan la presión diastólica por encima de 120 y la presión sistólica en menos de 100, son clara e intuitivamente 100% de individuos enfermos ya que desafían las leyes de la física cardíaca. Estos valores ya deberían corresponder a un edema pulmonar lo que conlleva a la muerte inmediata.

Relación percentilo de colesterol/índice de masa corporal por enfermo o no enfermo



PRESENCIA DE COLESTEROL EN RELACIÓN AL ÍNDICE DE MASA CORPORAL, VARIABLES CONTROLABLES CON LA DIETA, SON FACTORES INFLUYENTES EN EL DIAGNOSTICO

Este grafico contiene la relación entre el índice de masa corporal según cada categorización de colesterol siendo 1 normal, 2 poco normal y 3 muy poco normal.

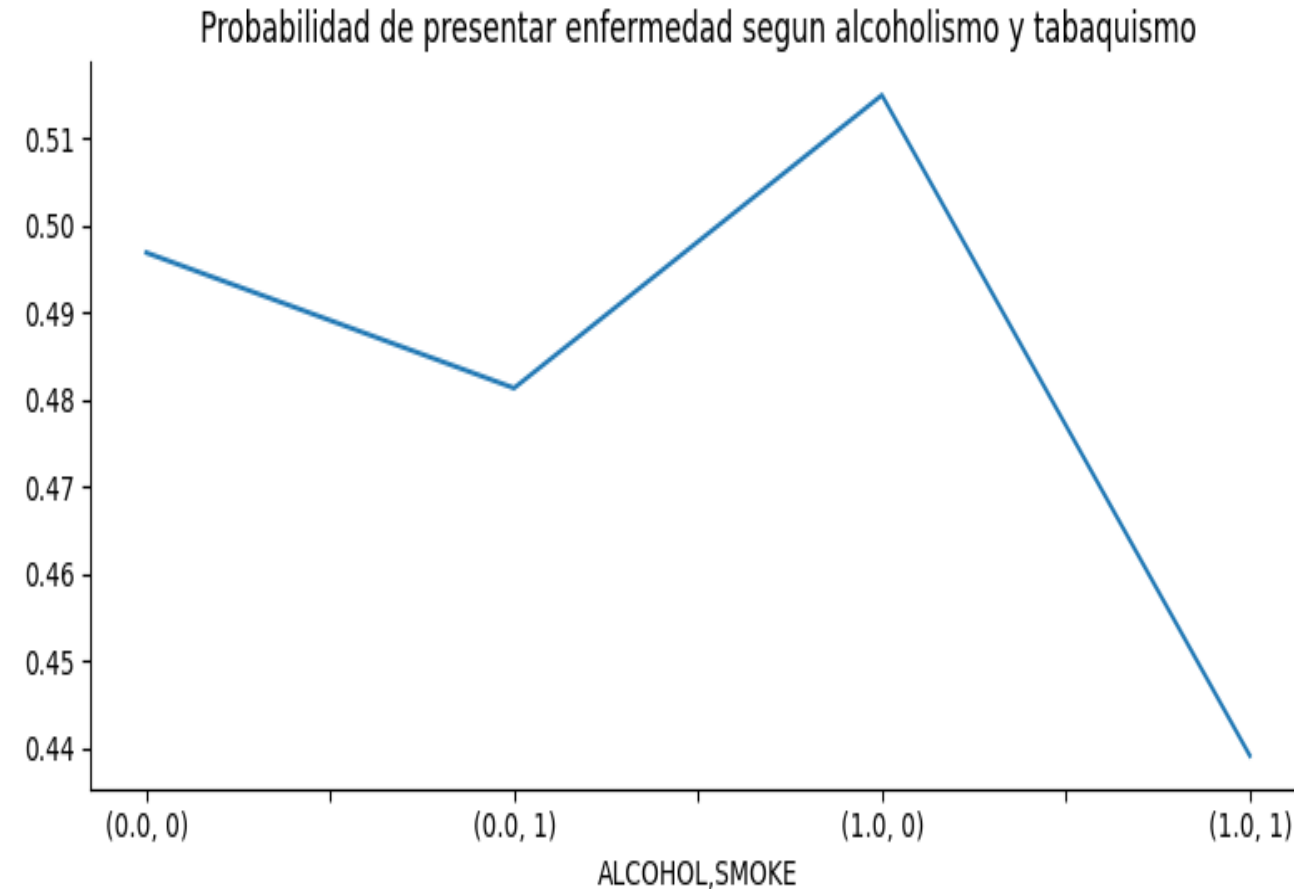
- Podemos dilucidar que los sujetos enfermos tienden presentar valores mayores de índice de masa corporal en las 3 valoraciones de colesterol.
- Mientras que los individuos obesos presentan enfermedad cardíaca en el 100% de los casos.
- Ambos parámetros son controlables con dieta y ejercicio.

EL ALCOHOLISMO Y EL TABAQUISMO SON FACTORES INFLUYENTES EN EL DIAGNOSTICO DE ENFERMEDAD.

En este gráfico encontramos datos de la probabilidad de ser un sujeto enfermo en relación al alcoholismo o actividad fumadora en conjunto.

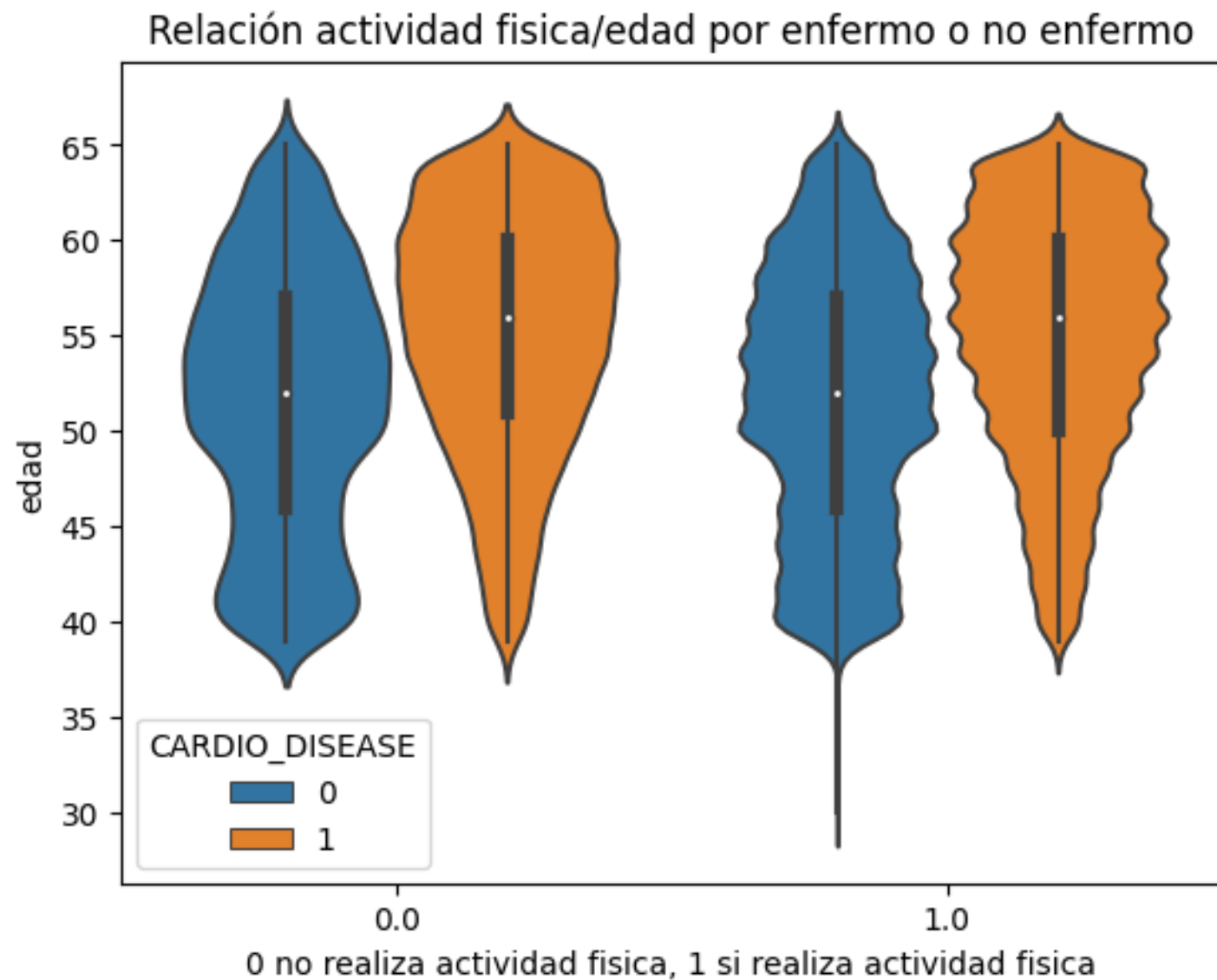
Podemos observar que los sujetos que no toman ni fuman tienen una probabilidad del 50% de presentar enfermedad cardíaca, mientras quienes sí fuman y no toman llegan a un 48%. Punto a favor de los fumadores.

En el caso de los bebedores no fumadores tienen una tendencia del 51% de desarrollar la enfermedad, mientras que los que conllevan ambos vicios presentan un 44%. No será el camino seguro hacia la salud pero al parecer tener ambos vicios mejora las posibilidades del sujeto.



LA ACTIVIDAD FÍSICA SEGÚN LA EDAD (AGE) IMPLICA MENOR INCIDENCIA DE ENFERMEDAD.

- En el grafico de violines a la derecha podemos ver que no hay una gran diferencia entre individuos que realizan o no actividad física pero sí una mayor incidencia de enfermedad a mayor edad.



INSIGHTS

INSIGHTS

- Presión sistólica por encima de 125
- Presión diastólica por sí solo no es un factor determinante.
- Presión sistólica como diastólica con valores por encima de lo normal.
- Presión diastólica por debajo de 80-100 y sistólica por encima 130.
- Presión diastólica por encima de 120 y presión sistólica en menos de 100.
- Valores altos de colesterol.
- Obesidad
- Edad elevada

MAYOR
PROBABILIDAD DE
ENFERMEDAD

INSIGHTS

- Presión diastólica por sí solo no es un factor determinante.
- Alcoholismo
- Tabaquismo
- Actividad física

BAJA
PROBABILIDAD O
NO
DETERMINANTE

SE APLICAN MODELOS DE MAGIC LEARNIG

Se evalúan todos los modelos y el que mejor indicadores de recall presenta en Naive Bayes Model con un 0.8 en predicciones de sujetos sanos y un 0.58 en enfermos.

Conclusión

Lamentablemente al ser tan bajo el porcentaje de acierto en sujetos enfermos y el 0,2 que predeciría como sanos estando enfermos son números poco aceptables para tomar este proyecto como un éxito.

Para el éxito del proyecto se necesita un dataset con mas indicativos y los mismos con una mayor relevancia.

- RANDOM FOREST
- SMV
- VECINOS CERCANOS
- NAIVE BAYES MODEL
- XGBOOST