

Detección de cardiopatías.

¿Cómo podemos ayudar a tomar decisiones?

Autor: Mauricio Flammini Zuccheri

Índice

1. Contexto del proyecto.
2. Preguntas de interés.
3. Análisis exploratorio de los datos.
4. Algoritmo
5. Conclusión.

Contexto

Es sabido que las enfermedades cardiovasculares son las que, probablemente, más muertes causan en todo el mundo. Se estima que aproximadamente un 30% de las muertes son a causa de este tipo de enfermedades, y por lo general sucede en personas menores de 70 años.

Audiencia: El principal motivo de este proyecto es poder brindar el análisis realizado a profesionales de la salud, para que éstos puedan utilizarlo como una herramienta y ser asistidos a la hora de tomar una decisión. De modo que a partir del trabajo propuesta, sea más sencilla la tarea de diagnosticar a un paciente.

Objetivo y preguntas de interés.

El principal objetivo de este proyecto es poder brindar una buena herramienta para ayudar a diagnosticar posibles cardiopatías a partir del machine learning.

De esta manera, podríamos preguntarnos cuál de los algoritmos de clasificación implementados es el más eficaz, y cuáles de las variables de la base de datos son las más importantes para el resultado.

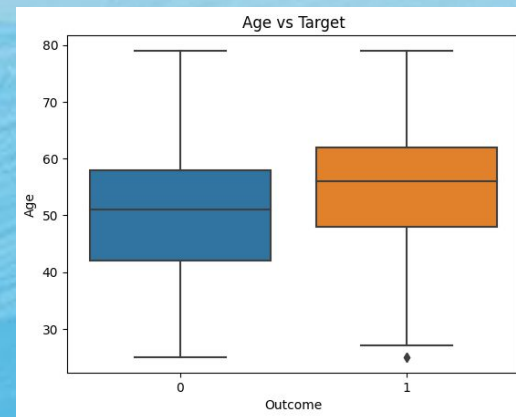
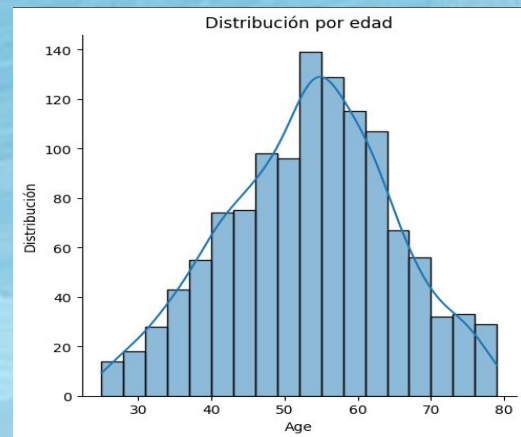
Análisis Exploratorio de Datos

Relación entre la edad de los pacientes y una cardiopatía diagnosticada

En la imagen de arriba a la derecha podemos ver cuántas personas hay dentro de un mismo rango etario para las más de 1200 personas que participaron del proyecto. La edad más frecuente entre los pacientes fue entre los 40 y 65 años.

La imagen de abajo nos dice que la mayoría de los pacientes sin cardiopatía (color azul) tienen entre 40 y 60 años.

En color naranja vemos que la mayor parte de pacientes con cardiopatías oscilan entre los casi 50 y 65 años

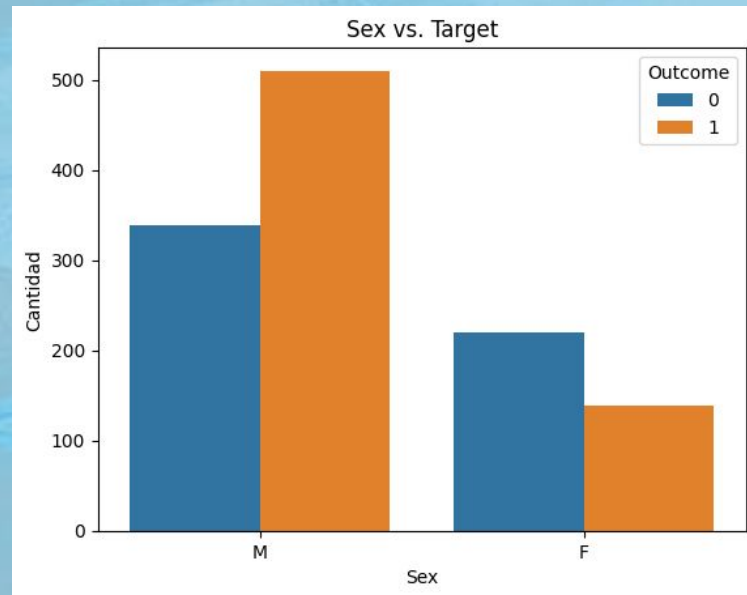


¿Es más habitual una cardiopatía en hombres o en mujeres?

En la imagen de la derecha primero podemos ver que en la muestra hubo más hombres que mujeres.

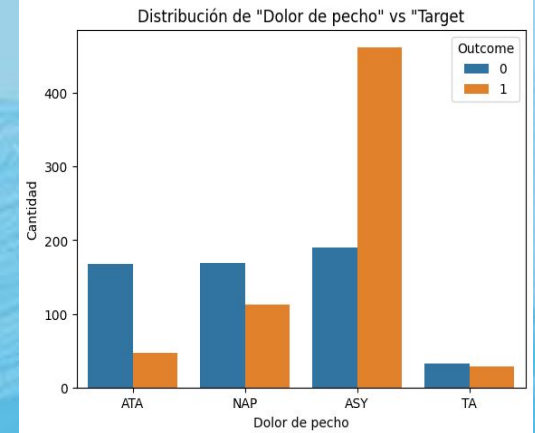
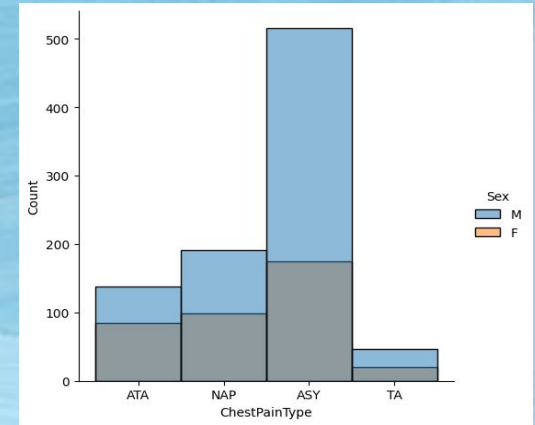
Luego podemos apreciar que para los hombres hubo más casos positivos que negativos.

En tanto a los mujeres, la distribución es casi la misma pero hubo más negativos



Tanto en hombres como mujeres, ¿predomina algún tipo de dolor de pecho sobre el resto?
¿si hay cardiopatía diagnosticada, algún tipo de dolor es más común?

Por lo que apreciamos en la imagen superior, claramente el tipo “asintomático” es el más frecuente. Es decir, la mayoría de los pacientes en el estudio no experimentaron dolor de pecho. En tanto a la relación entre tipo de dolor y cardiopatía: en los dolores de tipo anginoso, no anginoso y atípico, fue mayor la cantidad de casos sin cardiopatía. Los asintomáticos presentaron casi el doble de casos positivos de cardiopatía que negativos.

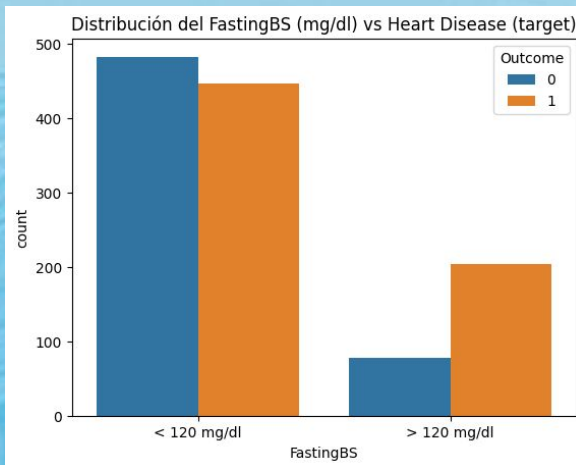
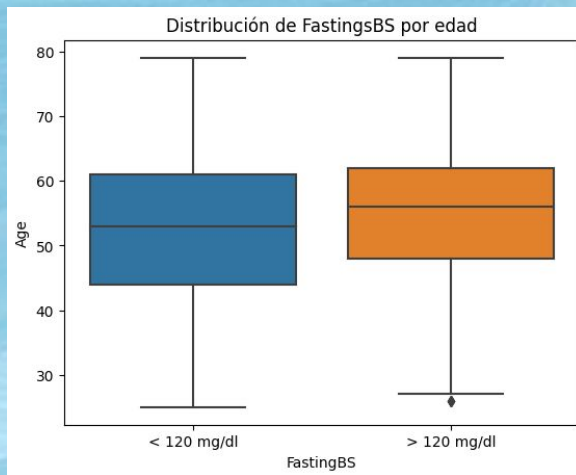


Relación entre una cardiopatía y los valores de glucemia.

Los valores de glucemia se han separado en mayor o menor a 120 mg/dl.

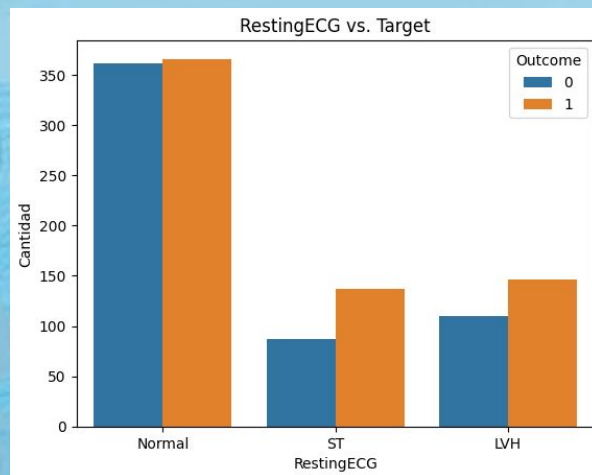
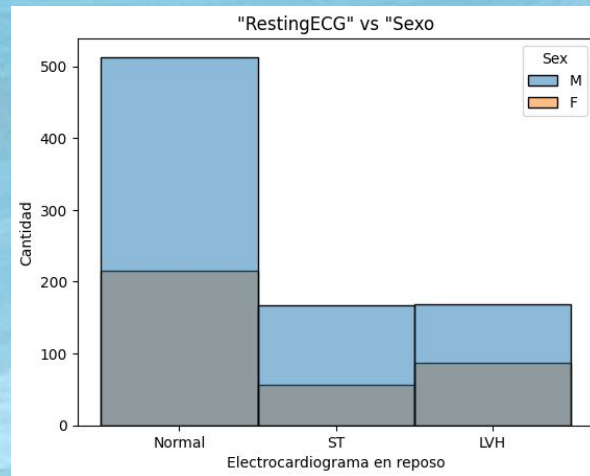
En el primer gráfico vemos que en los dos casos la mayoría de los pacientes están distribuidos en el mismo rango etario (entre 45 y 65 años aproximadamente, y un poco más acotado para >120 mg/dl).

En tanto a si tienen o no una cardiopatía según sus valores de glucemia: si es menor a 120 la cantidad de positivos y negativos es casi la misma. Para valores mayores a 120, la cantidad de positivos es prácticamente el doble a los negativos.



Características del electrocardiograma están más asociadas a una cardiopatía.

Un electrocardiograma sin anomalías predomina por sobre el resto. En tanto a la frecuencia de cada tipo de ECG con una posible cardiopatía, la cantidad de casos negativos y positivos es similar para la categoría “Normal”, pero para ST y LVH vemos que la cantidad de casos positivos es casi el doble a la de negativos.



Insights a partir de las imágenes y recomendaciones.

- La probabilidad de que una persona tenga una cardiopatía aumenta conforme aumenta la edad.
- El diagnóstico de una cardiopatía fue más frecuente en hombres que en mujeres.
- El tipo de dolor de pecho no parece ser un parámetro para diagnosticar. Puesto que en la mayoría de los casos fue “asintomático” y allí también había mayor diferencia entre positivos y negativos.
- Es menos frecuente que un paciente tenga glucemia mayor a 120 mg/dl, pero de ser así es mucho más probable una cardiopatía.
- La mayor diferencia de positivos y negativos en los ECG se dio para la anomalía ST.

Selección de algoritmos de clasificación.

En el proyecto se han explorando los siguientes modelos con algoritmos de clasificación:

- Árbol de decisión, Support Vector Classifier, K-Nearest Neighbors, Regresión Logística y AdaBoost.

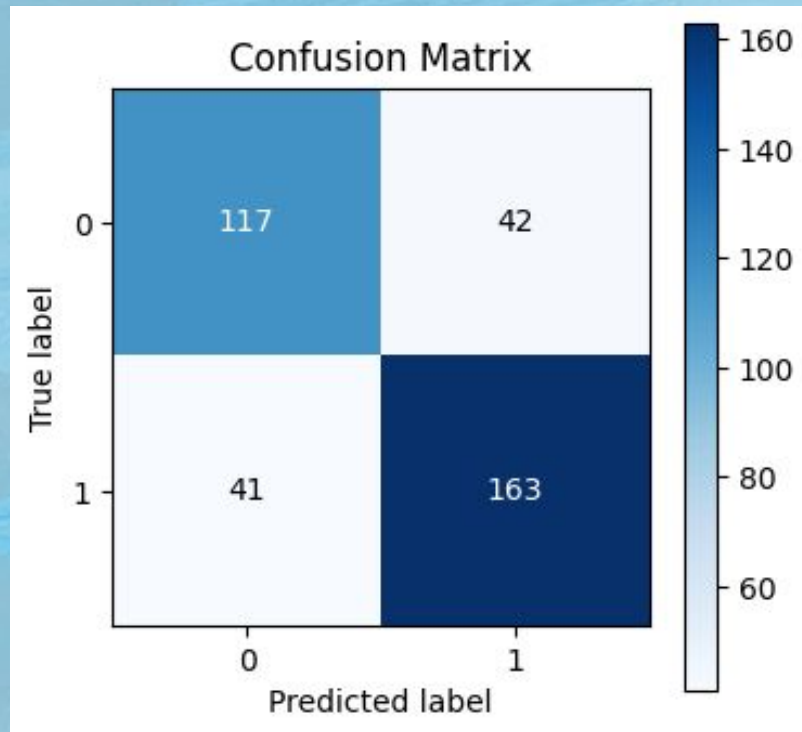
Para analizar el desempeño de cada modelo analizamos las siguientes métricas estadísticas: recall, precision y accuracy. Nos hemos centrado en recall.

El modelo que mejores resultados arrojó fue **Support Vector Classifier (SVM)**.

Support Vector Classifier (SVC).

- Recall: porcentaje de clasificación correcta de los verdaderos positivos. Fue del 80% en el conjunto de prueba.

Es decir, del total de 204 verdaderos positivos, predijo que eran positivos en 163 ocasiones, o sea, el 80%.



Importancia de cada variable.

Vemos una preponderancia de 'ST_Slope' por sobre las demás variables, luego 'ChestPainType' y 'Sex' fueron importantes para el modelo.

Estos resultados se condicen lo que podíamos intuir en la fase de Análisis Exploratorio de Datos.

