

ENFERMEDADES CARDIOVASCULARES

Proyecto de Machine Learning

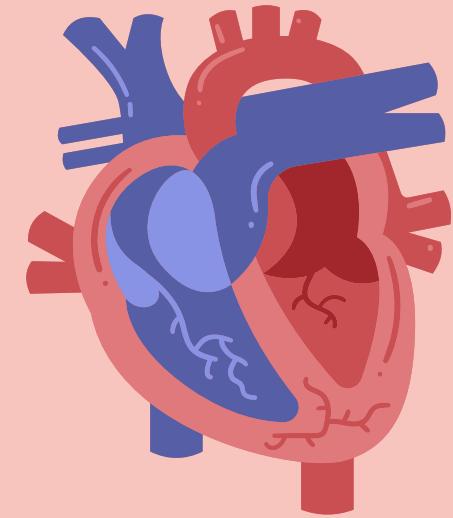
¿SE PUEDE PREDECIR UNA ENFERMEDAD CARDIOVASCULAR?

Las enfermedades cardiovasculares (ECV) continúan siendo una de las principales causas de mortalidad a nivel mundial.

El objetivo de este proyecto es explorar el conjunto de datos relacionado con las enfermedades cardíacas mediante un análisis exploratorio de datos (EDA) y aplicar algoritmos de clasificación para realizar predicciones.



¿CÓMO LO HAREMOS?

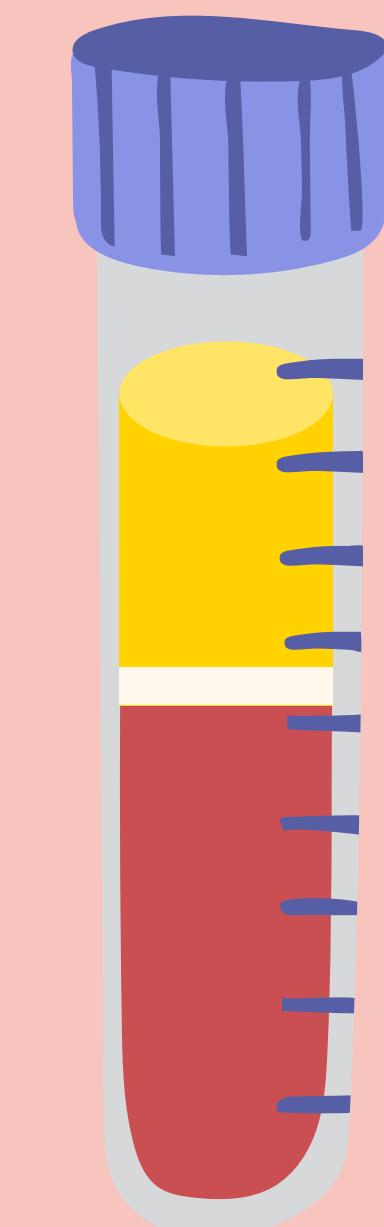


UCI Machine Learning Repository - Heart Disease Dataset

Este conjunto de datos fue creado para predecir la presencia de enfermedades cardíacas.

CONTENIDO

Dataset de 1025 filas



5 variables numérica

8 variables categóricas

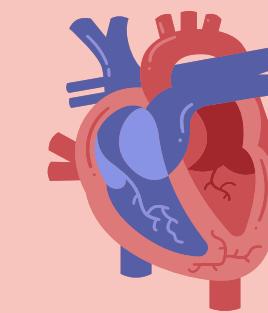
Target: binaria
0-1

PRIMERAS CONCLUSIONES

Pacientes a cierta edad sufren de esta enfermedad



Las mujeres tienen menos riesgo de padecer la enfermedad

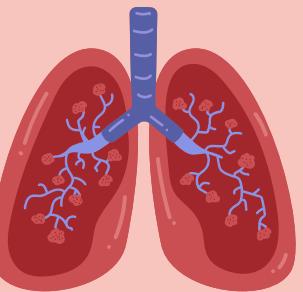


48.76% son pacientes con enfermedad

Frecuencia cardiaca

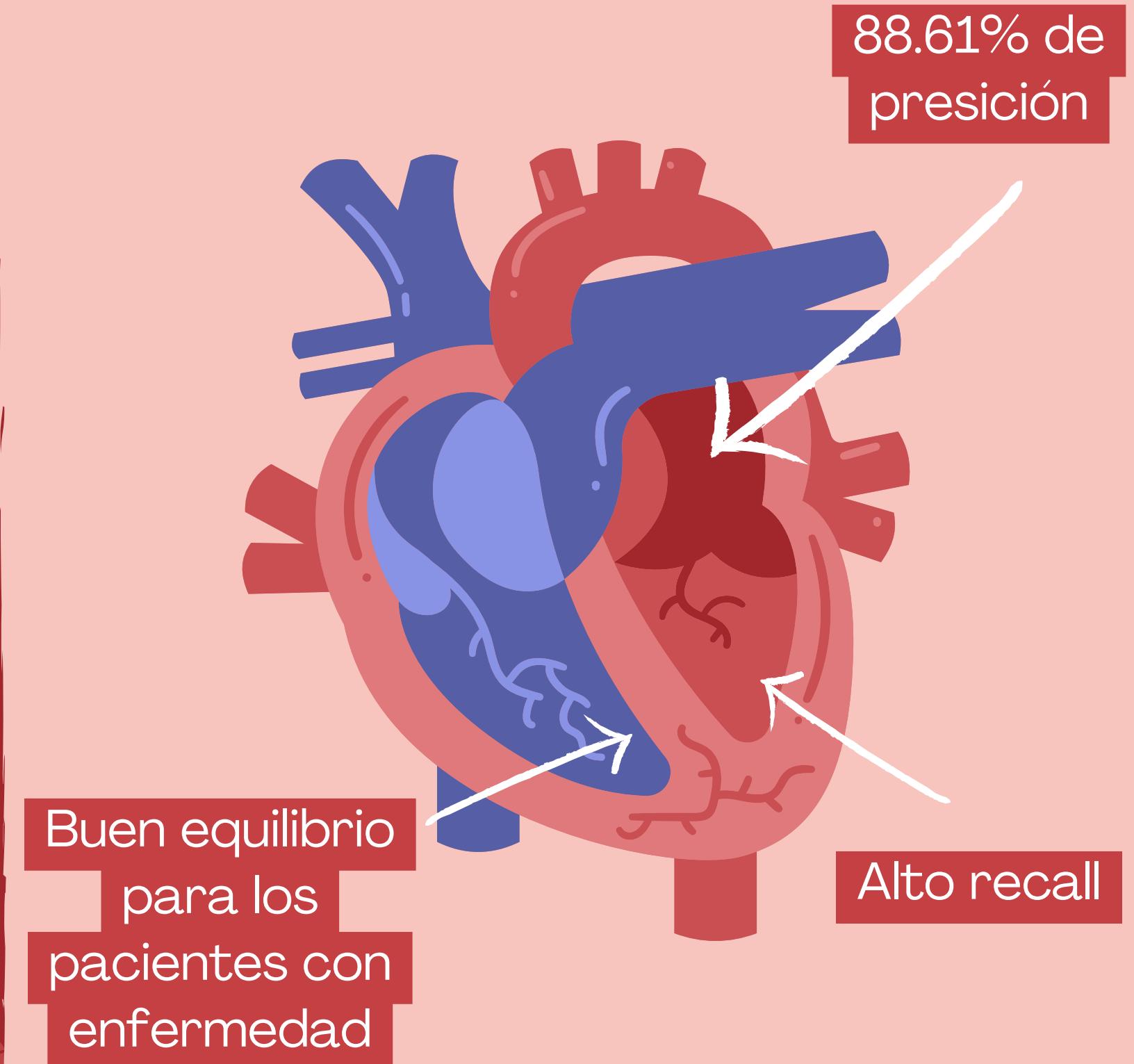
Dolor en el pecho

OldPeak

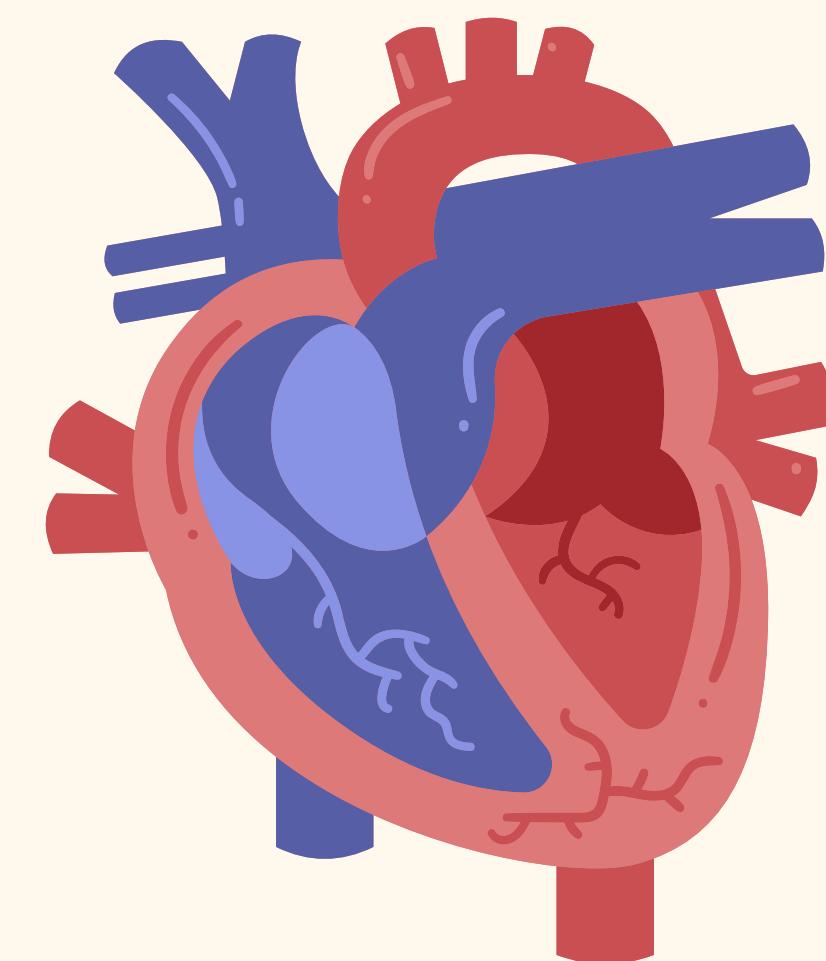


MODELO DE REGRESIÓN LOGÍSTICA

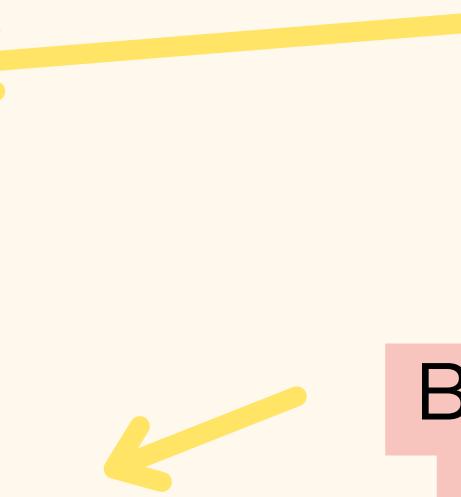
Modelo útil para detectar riesgos de enfermedad cardiovascular



NAIVES BAYES



Precisión del
85.64%



Buen balance entre
precisión y recall

RANDOM FOREST CLASSIFIER

Precisión del 95.54% predicciones correctas en la mayoría de los casos

Recall del 98%

Pocos pacientes sin la enfermedad son clasificados incorrectamente



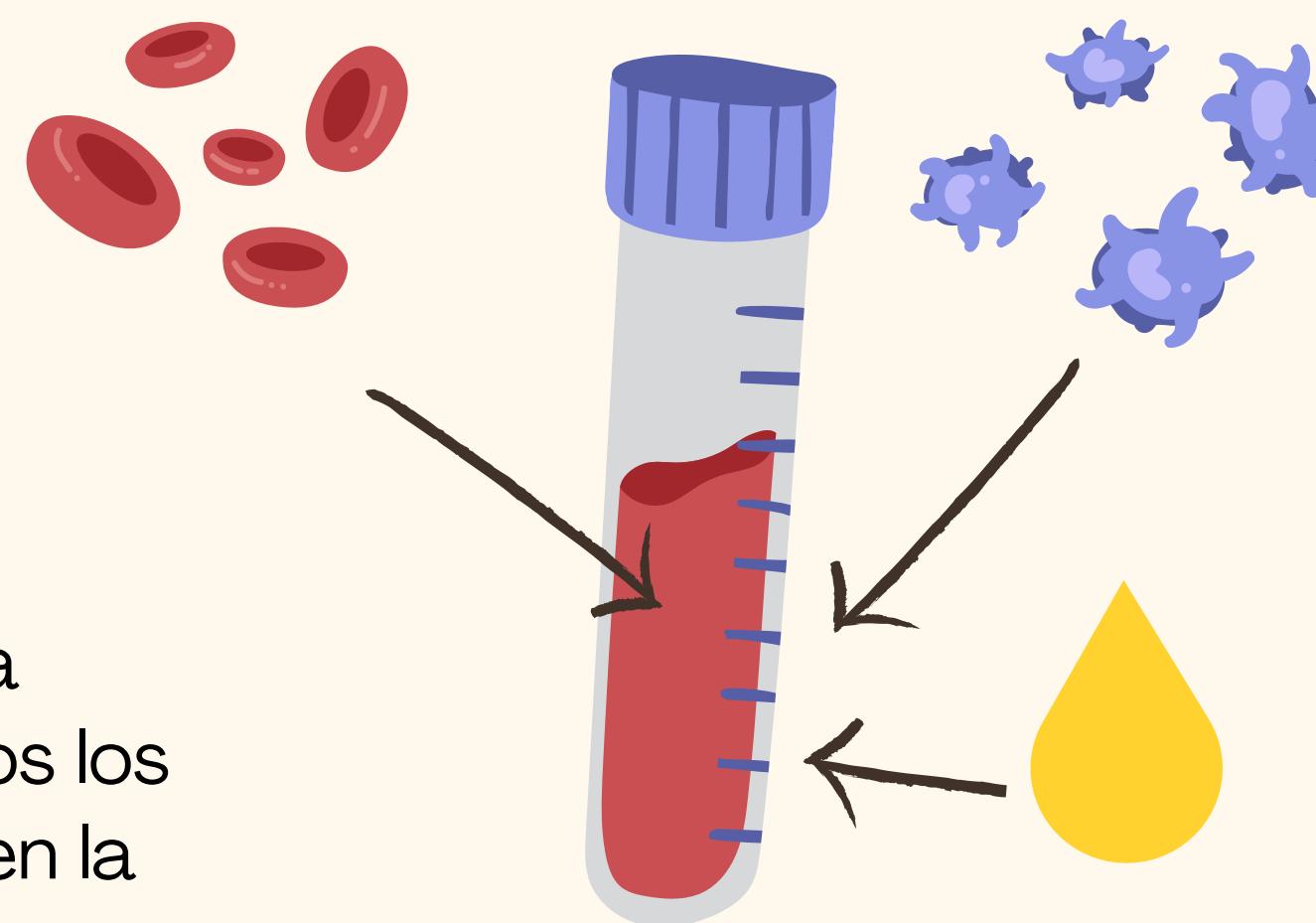
XGBOOST

Precisión

90%

Problemas:

El modelo no identifica correctamente a todos los pacientes que no tienen la enfermedad



Recall

98%

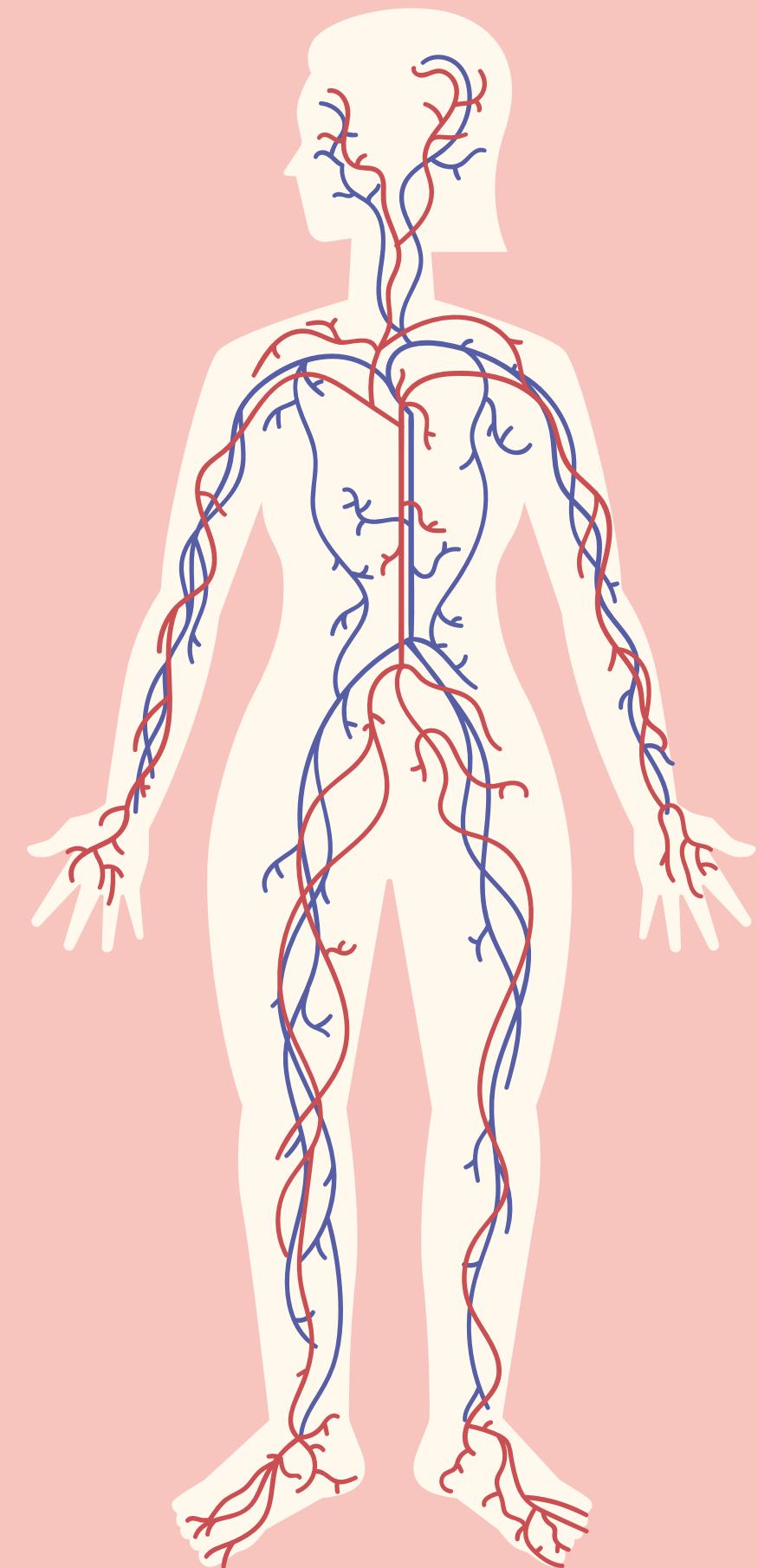
Beneficios

Mayor precisión en la predicción de pacientes sin enfermedad

ÁRBOLES DE DECISIÓN

El modelo muestra una gran capacidad para identificar correctamente tanto a pacientes con como sin enfermedad.

Los falsos positivos y falsos negativos son bajos, lo que muestra un buen equilibrio en la clasificación.



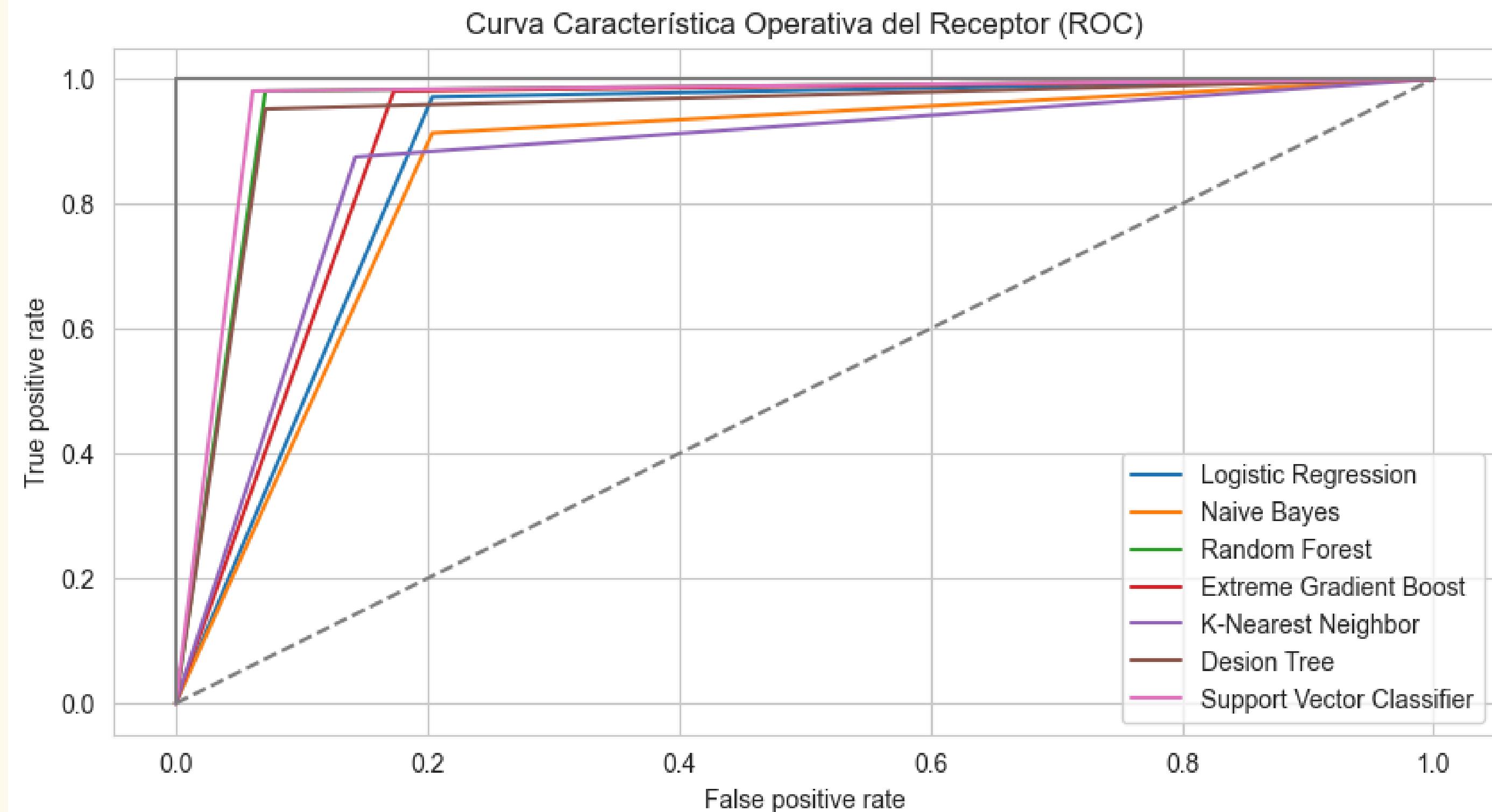
SUPPORT VECTOR CLASSIFIER

Rendimiento con una precisión global del 96%.

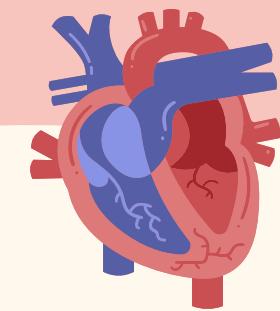
Modelo es muy efectivo para detectar a todos los pacientes, con o sin enfermedad



ROC

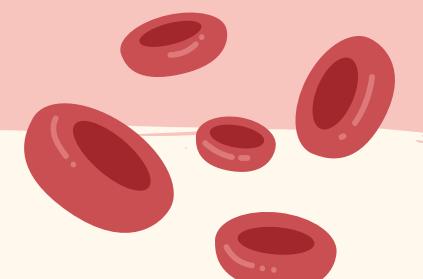


CONCLUSIÓN DE LOS MODELOS



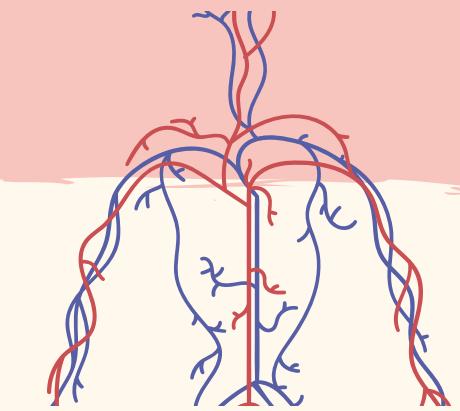
SVM

Mejor rendimiento con
una precisión del 96.04%



Random Forest

Rendimiento 95.54%



Naives Bayes

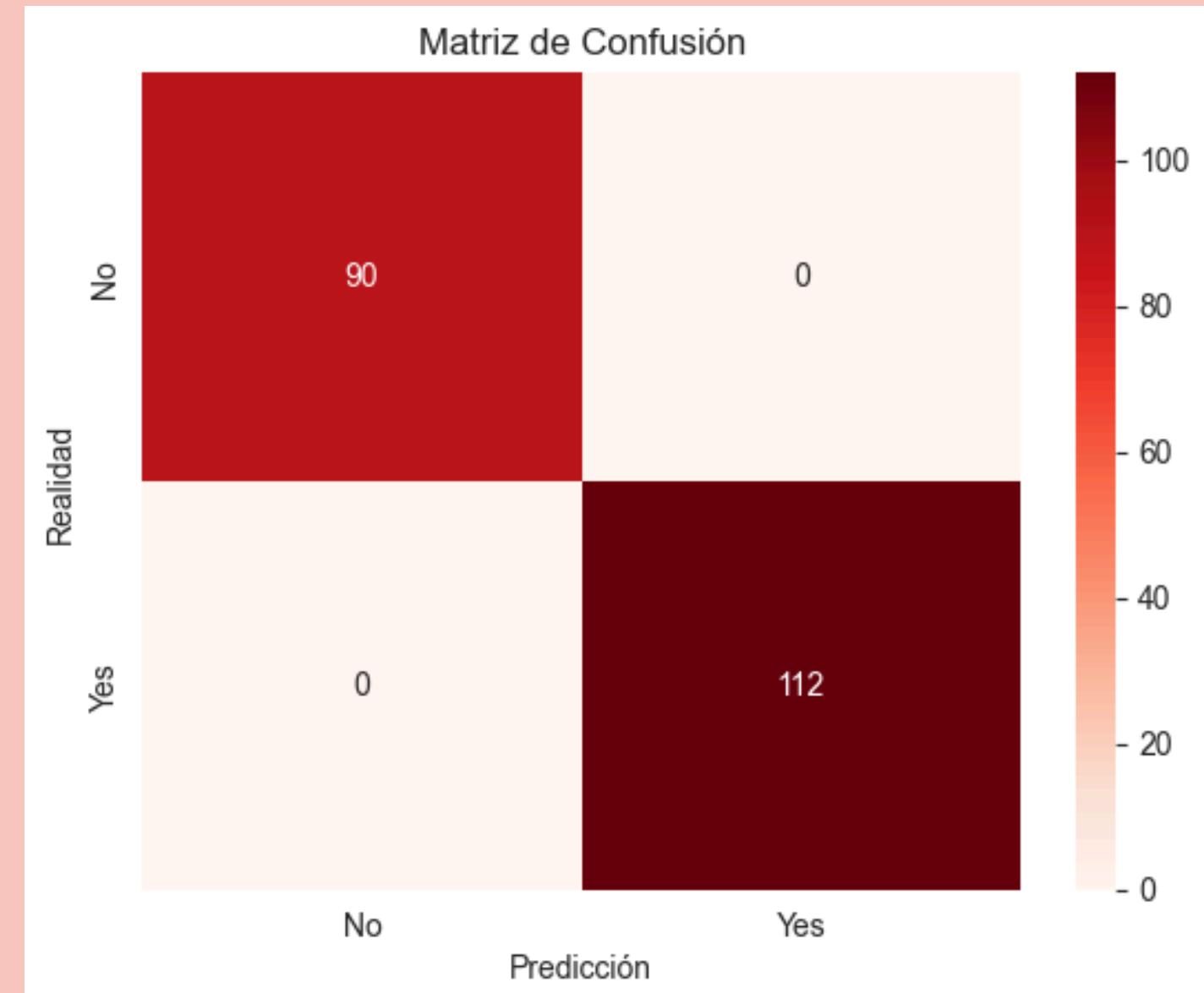
Peor rendimiento 85.64%

STACKING CLASSIFIER

El Stacking Classifier entrenado ha obtenido una precisión del 100% en el conjunto de prueba, lo que indica que ha clasificado todas las muestras correctamente.

Está ofreciendo una excelente capacidad predictiva.

El modelo también tiene un rendimiento excelente en el conjunto de entrenamiento.



CONCLUSIÓN

Este estudio demuestra la importancia del análisis de datos en el campo de la salud.

Los modelos de predicción pueden ser herramientas clave para la detección temprana de enfermedades cardíacas y podrían ayudar a los profesionales de la salud a tomar decisiones más informadas.





Gracias por su atención

Carolina Monzón Felipe