



Predicción de enfermedades cardíacas

Daniel Arias

Universidad San Francisco de Quito

12-04-2025



Objetivo del proyecto

Predecir la presencia de enfermedades cardíacas y clasificarlas como leve, moderada, grave o muy grave.

Relevancia Estratégica

Clasificar enfermedades cardíacas es clave para anticipar riesgos, personalizar tratamientos y salvar vidas. Este proyecto no solo mejorará la atención en el sector médico, sino que optimiza recursos y eleva la eficiencia organizacional.

Contexto y Alcance



Contexto y Alcance

CONTEXTO

En Ecuador, pacientes pueden esperar meses para una cita especializada en el sector público, con tiempos promedio de hasta **90 días** para consultas cardiológicas, lo que retrasa diagnósticos críticos y pone vidas en riesgo.

ALCANCE

Selección de un modelo de Machine Learning para la clasificación de enfermedades cardíacas.

Entendimiento de datos



303 Entradas

3 Valores
faltantes

0 Entradas
duplicadas

Los valores
faltantes
únicamente
representan el
**1% de los
datos**



13 Características:

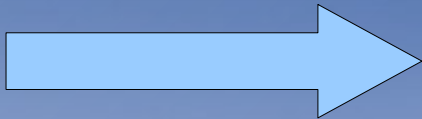
- Edad: Años del paciente.
- Sexo: Masculino o femenino.
- Dolor torácico (cp): Tipo de dolor en el pecho.
- Presión arterial (trestbps): En reposo, mm Hg.
- Colesterol (chol): Nivel en sangre, mg/dl.
- Azúcar en ayunas (fbs): Nivel elevado o no.
- Electrocardiograma (restecg): Resultados en reposo.
- Frecuencia cardíaca (thalach): Máxima durante ejercicio.
- Angina por ejercicio (exang): Presencia o ausencia.
- Depresión ST (oldpeak): Cambio por ejercicio.
- Pendiente ST (slope): Forma del segmento ST.
- Vasos principales (ca): Número afectados.
- Talio (thal): Resultados de prueba de flujo sanguíneo.

Preparación de datos

Se eliminaron entradas con datos faltantes

Se reemplazaron los nombres de las variables
categóricas

Cp	fbs
1	0
4	1



chest_pain_type	fasting_blood_sugar_g_120
Typical angina	True
Asymptomatic	False

Convertir a valores booleanos las entradas de texto

Modelado

Modelos a probar

- AdaBoost
- RandomForest
- Bagging
- XGBoost

Métricas a tomar en cuenta

- Precisión:
¿Qué pacientes de los que predije dentro de una clase realmente pertenecen ahí?

Metodología de entrenamiento y validación

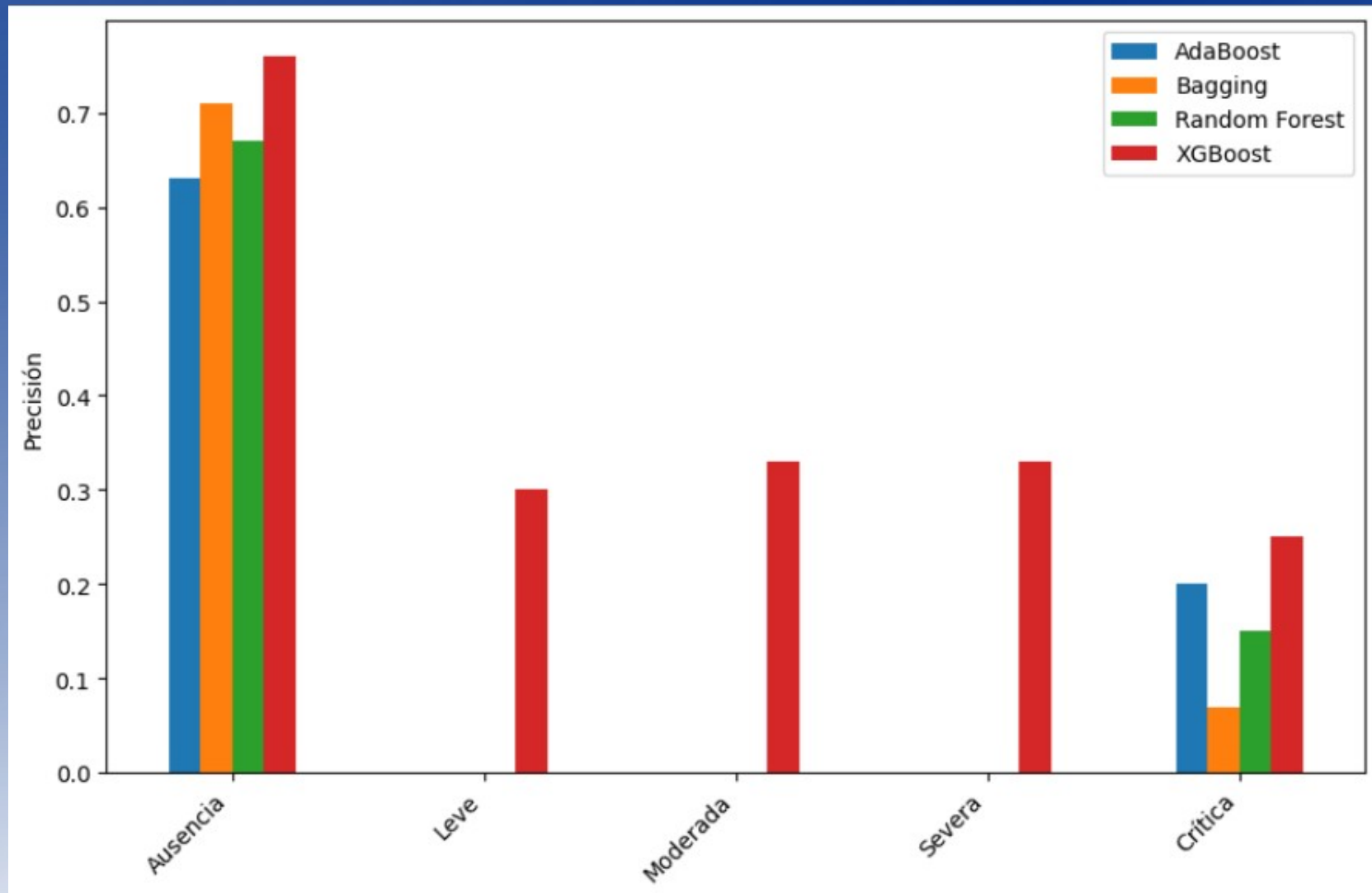
• Selección del modelo

- Entrenamiento con una fracción del total de los datos (75%).
- Evaluación de resultados con la fracción restante.

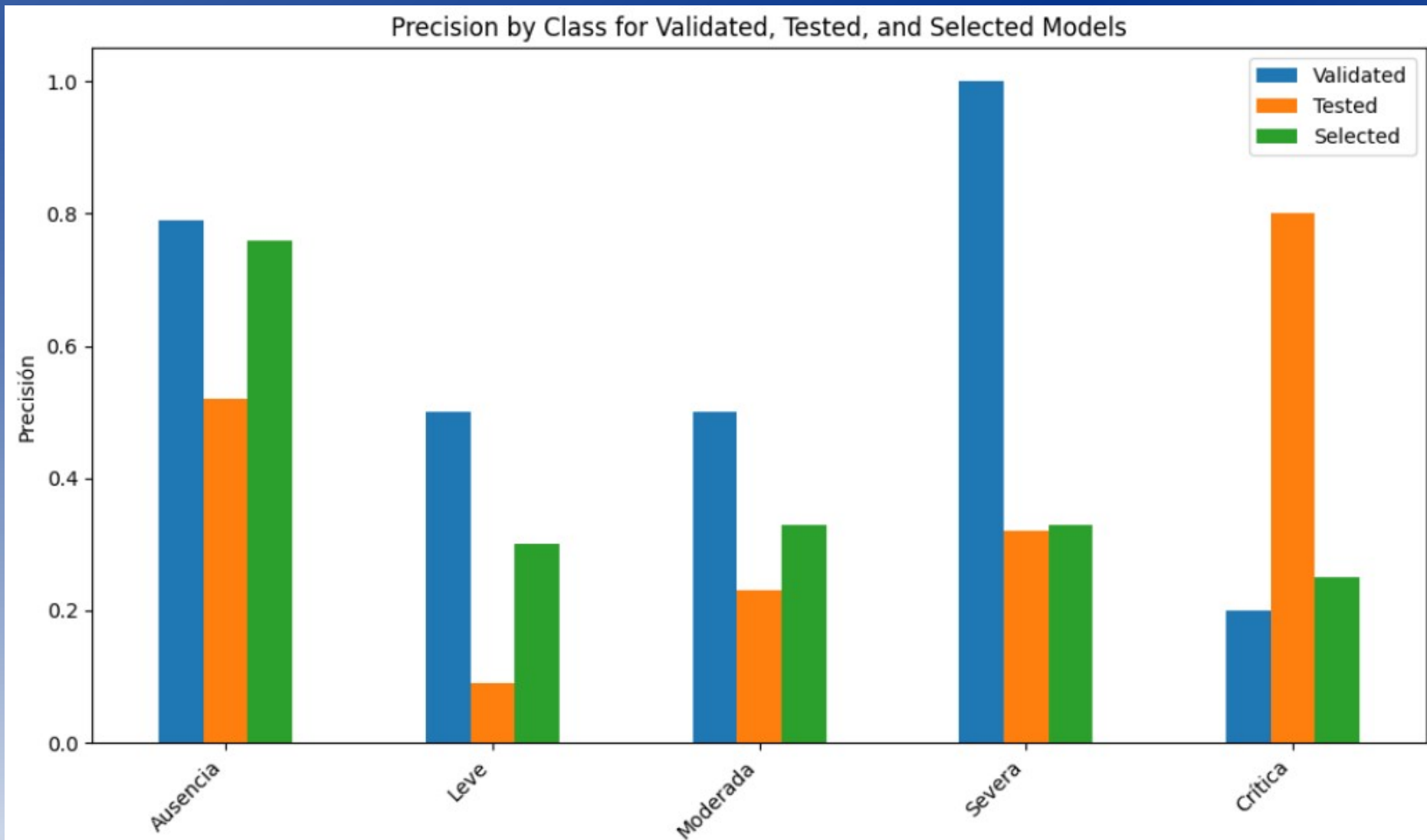
• Ajuste de Hiperparámetros.

- Entrenamiento y validación con una fracción de los datos originales.
- Evaluación de resultados con datos de prueba (25% de los datos originales)

Resultados: Selección del modelo



Resultados: Tuneo del modelo XGBoost



Conclusiones

1

El modelo logra predecir **8 de cada 10** casos de enfermedades cardíacas críticas.

2

Es necesario pulir el Modelo para incrementar precisión para la detección de todas las clases involucradas en el problema.

Plan de implementación y Próximos pasos

1

Optimización de hiper-parámetros

2

Validación Final

3

Integración del modelo

4

Capacitación del personal

5

Monitoreo y mantenimiento

Muchas gracias por su atención

