

# Construcción de Scores Crediticios con Modelos de Aprendizaje Automático

## Curso Herramientas de Datos 2

Joseph Romero Chinchilla, Sebastián Astúa Morales,  
Pablo Alonso Chaves Madrigal

Universidad de Costa Rica  
Departamento de Ciencias Actuariales y Matemáticas Puras

Julio, 2025

# Agenda

- 1 Introducción
- 2 Metodología
- 3 Modelos
- 4 Resultados
- 5 Conclusiones

# ¿Qué es un Score Crediticio?

- Evalúa la solvencia de solicitantes de crédito.
- Variables: historial de pagos, uso de crédito, antigüedad crediticia.
- Apoya decisiones sobre aprobación y condiciones del préstamo.

- Modelos como Random Forest, Gradient Boosting y XGBoost han mejorado la predicción de impago.
- Permiten capturar relaciones complejas y manejar desbalanceo.

- Dataset de Kaggle: *Credit Risk Dataset*.
- Codificación one-hot, normalización y SMOTE.
- Evitamos *data leakage* removiendo variables como `historial_impago`.

- pandas, scikit-learn, xgboost, imblearn.
- Visualización: matplotlib, seaborn.
- Balanceo de clases con `compute_sample_weight` y SMOTE.

- Ensamble de árboles entrenados en subconjuntos aleatorios.
- Votación mayoritaria.

# Gradient Boosting

- Construcción secuencial de árboles.
- Cada nuevo árbol corrige errores del anterior.

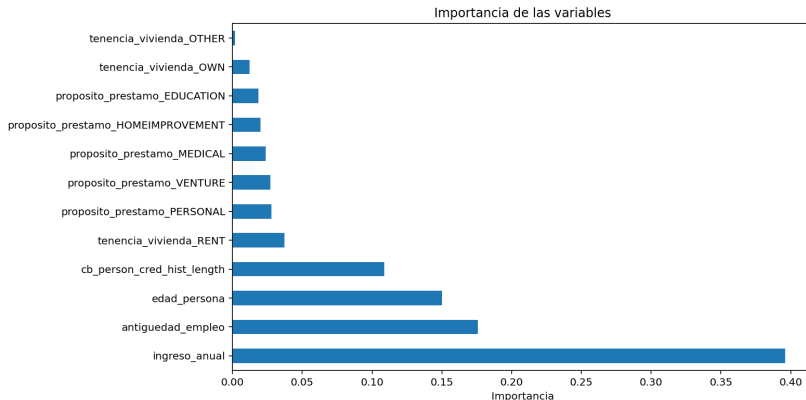


- Versión optimizada de Gradient Boosting.
- Regularización, manejo automático de valores faltantes.
- Mejor desempeño en datos financieros desbalanceados.

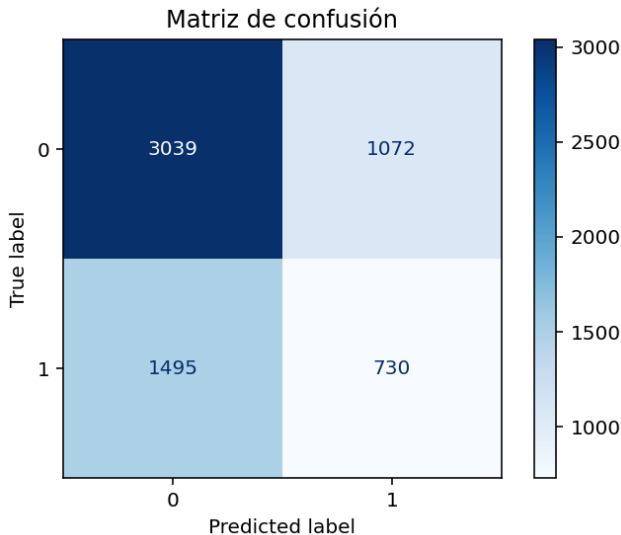
# Comparación de Desempeño

| Métrica   | RF     | GB     | XGB           |
|-----------|--------|--------|---------------|
| Accuracy  | 0.5949 | 0.6214 | <b>0.6373</b> |
| Precision | 0.5772 | 0.5838 | <b>0.6031</b> |
| Recall    | 0.5949 | 0.6214 | <b>0.6373</b> |
| F1-Score  | 0.5835 | 0.5871 | <b>0.6019</b> |

# Importancia de Variables



# Matriz de Confusión XGBoost



- XGBoost fue el modelo con mejor rendimiento.
- Aporta alto valor para identificar clientes de riesgo.
- Balance adecuado entre precisión y sensibilidad.

- Continuar refinando modelos con más datos.
- Evaluar técnicas para reducir falsos negativos.
- Explorar interpretabilidad y fairness en modelos.