

Iniciativa del Banco Omega: Optimización de la Morosidad en Pagos con Random Forest e Inteligencia Artificial

Carlos Paredes
Esteban Osejo

Resumen— La iniciativa del Banco Omega se concentró en examinar y resolver el aumento de la morosidad en los pagos con tarjetas de crédito, que se tradujo en menores ganancias. Los datos de los consumidores, que incluían historial crediticio, patrones de gasto e información personal, se procesaron mediante técnicas de análisis de datos y aprendizaje automático. El objetivo principal fue determinar las variables que afectan la morosidad y desarrollar un modelo de clasificación para pronosticar si un cliente realizará los pagos de su préstamo a tiempo. Los resultados muestran que, en exactitud, Random Forest sintonizado logra una exactitud de 81.27 %, en precisión 67.37 %, en sensibilidad 28.87 % y en especificidad 96.05 %. Destacando parámetros ligeramente superiores en comparación con XGBoost. Subrayando la importancia de seleccionar el algoritmo adecuado basado en las características del conjunto de datos y los objetos de clasificación.

Abstract— The Omega Bank initiative focused on examining and addressing the increased delay in credit card payments, which resulted in lower earnings. Consumer data, including credit history, spending patterns and personal information, were processed using data analysis and machine learning techniques. The main objective was to identify the variables that affect default and to develop a classification model to predict whether a client will pay his loan on time. The results show that, in accuracy, Random Forest tuned achieves a precision of 81.27%, precision 67.37%, sensitivity 28.87% and specificity 96.05%. Highlighting slightly higher parameters compared to XGBoost. Underlining the importance of selecting the appropriate algorithm based on the characteristics of the data set and the classification objects.

I. INTRODUCCIÓN

Para maximizar la precisión de la categorización de datos, esta investigación compara y optimiza dos técnicas de aprendizaje automático. XGBoost sintonizado y Random Forest sintonizado.

Omega Bank está llevando a cabo una iniciativa de análisis de datos y aprendizaje automático en respuesta al aumento de los atrasos en los pagos con tarjetas de crédito y su efecto en la rentabilidad.

El objetivo es identificar las variables que afectan la morosidad de los consumidores y crear un modelo de pronóstico para determinar si un cliente realizará los pagos de su préstamo a tiempo.

Para mejorar la gestión del riesgo crediticio y fortalecer las conexiones con los clientes, este proyecto procesa datos de los consumidores, incluida información personal, historial crediticio y hábitos de gasto, utilizando técnicas sofisticadas como Bagging y Boosting.

II. METODOLOGÍA

Se utiliza un procedimiento exhaustivo para abordar el análisis de la morosidad en los pagos con tarjeta de crédito y comienza con la recopilación de datos históricos del cliente, que incluyen historial crediticio, patrones de gastos, información personal, pagos anteriores y otra información pertinente.

El siguiente paso del proceso es el preprocesamiento de datos, que implica la limpieza de datos para eliminar errores, datos faltantes y valores atípicos y garantizar la calidad de los datos que se utilizarán. Además, utilizando métodos como el etiquetado ordinal o la codificación one-hot, los datos categóricos se convierten en variables numéricas para su posterior análisis.

BAGGING

El uso de Random Forest para implementar el embolsado implica entrenar numerosos árboles de decisión con subconjuntos aleatorios de datos y luego fusionar los resultados para disminuir el sobreajuste y aumentar la precisión como se observa en la siguiente Figura 1.

Entrenamiento y Evaluación del Modelo Random Forest

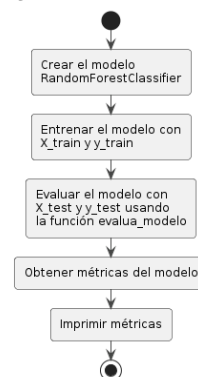


Fig. 1. Diagrama de flujo Random Forest [1]

En la siguiente Figura 2 se puede observar el diagrama de flujo de como es el funcionamiento de Random Forest sintonizado.

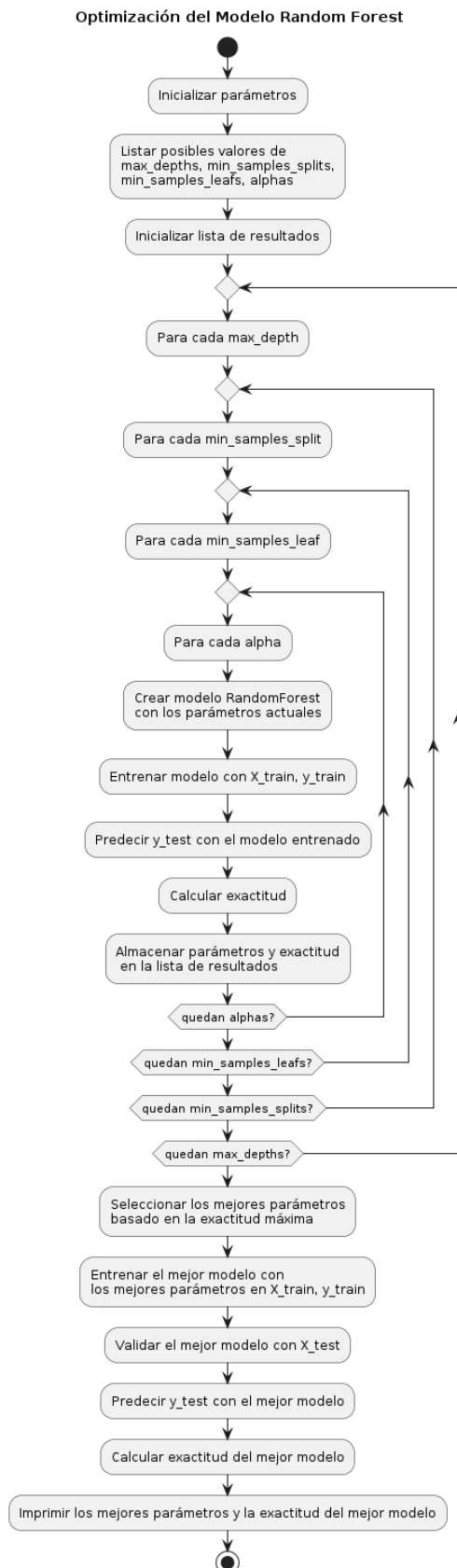


Fig. 2. Diagrama de flujo Random Forest sintonizado

BOOSTING

XGBoost sigue un enfoque iterativo, entrenando una serie de modelos uno tras otro. Cada nuevo modelo se enfoca en corregir los errores cometidos por el modelo anterior, lo que resulta en una mejora constante de la precisión del conjunto con el tiempo como se observa en la siguiente Figura 3.

Entrenamiento y Evaluación del Modelo XGBoost

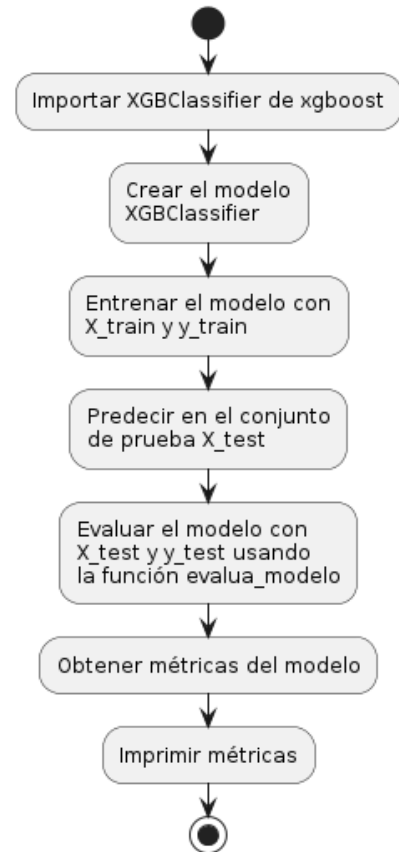


Fig. 3. Diagrama de flujo XGBoost

En la siguiente Figura 4 se puede observar el modelo sintonizado de XGBoost y su funcionamiento.

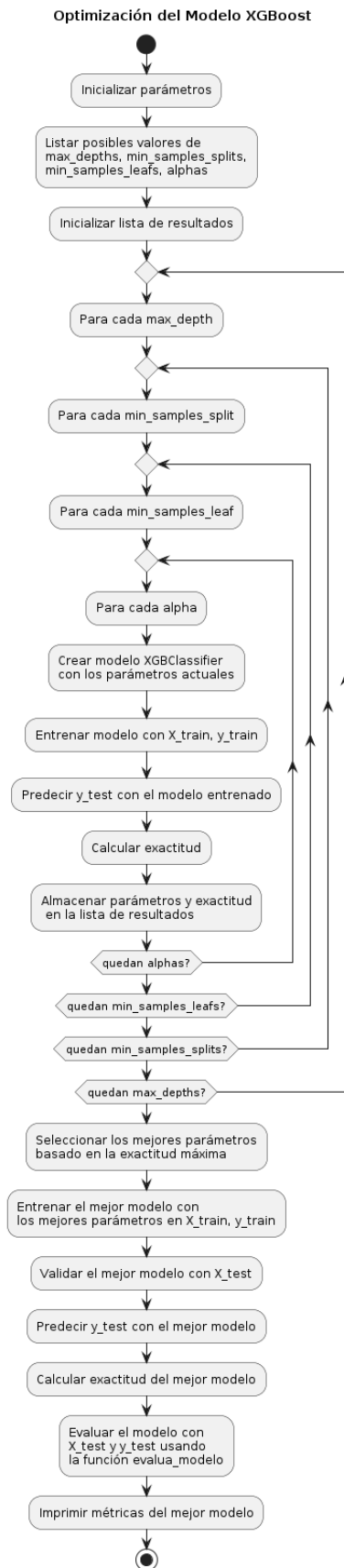


Fig. 4. Diagrama de flujo XGBoost sintonizado

III. RESULTADOS

En este estudio se compararon cuatro modelos de aprendizaje automático: XGBoost, Random Forest, XGBoost sintonizado y Random Forest sintonizado. Se utilizó un conjunto de datos específico para evaluar estos modelos con el fin de identificar cuál funciona mejor en términos de eficiencia y precisión para la clasificación de datos.

El objetivo principal de esta comparación es determinar qué modelo es mejor para la tarea de clasificación dada, teniendo en cuenta factores como los tiempos de entrenamiento y predicción, la interpretabilidad de las reglas producidas y la precisión de la predicción. Como se observa en la Tabla I.

Métricas	XGBoost SINTONIZADO	Random Forest SINTONIZADO
Accuracy	0.812444	0.812667
Precision	0.672963	0.673734
Sensitivity	0.287733	0.288743
Specificity	0.960536	0.960536
F1	0.403112	0.404240

Tabla I. Comparación de resultados

IV. CONCLUSIONES

La exactitud obtenida es de 81.244% para XGBoost sintonizado y 81.2667% para Random Forest sintonizado, lo que indica que ambos modelos mostraron una eficacia muy similar en este estudio. En términos de precisión, XGBoost sintonizado alcanzó un 67.296% y Random Forest sintonizado un 67.3734%, mostrando un rendimiento comparable.

Sin embargo, en términos de sensibilidad, XGBoost sintonizado presentó una menor precisión con un 28.773% en comparación con el 28.8743% de Random Forest sintonizado. Ambos modelos mostraron una especificidad alta, con XGBoost sintonizado y Random Forest sintonizado obteniendo un 96.0536%. La puntuación F1, que equilibra la precisión y la sensibilidad, fue de 40.3112% para XGBoost sintonizado y 40.4240% para Random Forest sintonizado.

V. REFERENCIAS

- [1] J. Brownlee, "Bagging and Random Forest Ensemble Algorithms for Machine Learning," MachineLearningMastery.com. [Online]. Disponible: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/>. [Obtenido: 14-May-2024].
- [2] J. Brownlee, "XGBoost Algorithm in Machine Learning," Analytics Vidhya. [Online]. Disponible: <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>. [Obtenido: 14-May-2024]