



UNIVERSIDAD INTERNACIONAL DEL ECUADOR

FACULTY OF TECHNICAL SCIENCES

SCHOOL OF MECHATRONICS ENGINEERING

COMPARATIVA ENTRE ALGORITMOS DE CLASIFICACIÓN DEL BANCO OMEGA

- Proyecto De Endomorfismo -

Mireya Tacuri

Comparativa entre algoritmos de clasificación del Banco Omega

Mireya Tacuri^{1,*},

¹School of Mechatronic Engineering, International University of Ecuador, Quito, Ecuador

*Autor Principal/Corresponding author, e-mail: mitacuriar@uide.edu.ec

Abstract—

Omega Bank's initiative focused on investigating and addressing the rise in credit card delinquencies, which has led to lower profits. Use data analytics and machine learning techniques to process consumer data, including credit records, spending patterns and personal information. Identify variables that influence delinquency and develop ranking models to predict whether customers will pay on time or in installments. We have the following characteristics: credit limit, gender, education, marital status, age, payment history, bill amount, previous payments, and whether the next payment is collected late.

Keywords—data analysis, bagging, boosting, delinquency, on-time payment.

Resumen—

La iniciativa del Banco Omega se concentró en investigar y abordar el aumento de la morosidad en las tarjetas de crédito, que ha provocado menores beneficios. Utilice análisis de datos y técnicas de aprendizaje automático para procesar datos de los consumidores, incluidos registros de crédito, patrones de gasto e información personal. Identificar variables que influyen en la morosidad y desarrollar modelos de clasificación para predecir si los clientes pagarán a tiempo o en cuotas. Contamos con las siguientes características: límite de crédito, género, educación, estado civil, edad, historial de pagos, monto de la factura, pagos anteriores y si se cobra el siguiente pago con retraso.

Palabras Clave—análisis de datos, bagging, boosting, morosidad, pago a tiempo.

I. OBJETIVOS

- Identificar las variables que afectan la morosidad de los consumidores y crear un modelo de pronóstico para determinar si el cliente realizará los pagos de su préstamo a tiempo.

II. INTRODUCCION

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

A. Bagging

Bagging es una técnica que implica la combinación de múltiples modelos de aprendizaje para mejorar la precisión general del modelo. Funciona construyendo múltiples modelos independientes a partir de diferentes subconjuntos de datos de entrenamiento. Estos subconjuntos se generan mediante un muestreo aleatorio con reemplazo, lo que significa que una instancia de datos puede aparecer en múltiples subconjuntos. Cada modelo se entrena de forma independiente con uno de estos subconjuntos y luego las predicciones de todos los modelos se combinan para obtener una predicción final, ya sea por votación (en problemas de clasificación) o promediando

(en problemas de regresión). Ejemplos comunes de algoritmos de Bagging incluyen Random Forest y Bagged Decision Trees.

B. Boosting

Boosting es una técnica que también combina múltiples modelos de aprendizaje, pero a diferencia del Bagging, los modelos se construyen secuencialmente, donde cada modelo intenta corregir los errores del modelo anterior. En cada iteración, el algoritmo asigna pesos a las instancias de datos, dándole más peso a las instancias que fueron mal clasificadas en las iteraciones anteriores. Esto permite que el modelo se centre en las instancias difíciles de clasificar y mejore gradualmente su rendimiento. Los modelos individuales en Boosting suelen ser modelos débiles, como árboles de decisión poco profundos. Ejemplos populares de algoritmos de Boosting incluyen AdaBoost, Gradient Boosting, y XGBoost

III. METODOLOGIA

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.

Primero se realizan los ajustes de hiperparámetros y la validación cruzada. Esto implica el uso de métodos como la validación cruzada k-fold, que divide los datos en conjuntos de entrenamiento y prueba repetidamente para evaluar la capacidad del modelo para generalizar. De manera similar, métodos como Grid Search se utilizan para el ajuste de hiperparámetros, que busca la mejor combinación de parámetros posible para mejorar el rendimiento del modelo.

Luego, el modelo se evalúa utilizando métricas que incluyen la matriz de confusión, el área bajo la curva ROC (AUC-ROC), la puntuación F1, la precisión y la recuperación. Estas medidas brindan una imagen integral del desempeño del modelo y muestran qué tan bien predice la morosidad en los pagos. Para determinar las variables más importantes en la predicción de la morosidad, también se realiza un análisis de importancia de las características. Esto facilita una comprensión más profunda de los elementos que influyen en la gestión del riesgo crediticio y el comportamiento financiero de los clientes.

En última instancia, después de someterse a validación y evaluación, el modelo se implementa en un entorno del mundo real para generar predicciones en tiempo real. Para garantizar la eficacia del modelo en el tiempo, se establece un método de seguimiento continuo para evaluar su rendimiento en escenarios reales y realizar las mejoras necesarias.

IV. RESULTADOS

Los resultados dados nos dan una exactitud de 0.82 ya sea en RandomForest o en XGBoots en el cual tenemos nuestras Exactitud:0.82, Precision:0.67, Sensibilidad:0.36, Especificidad:0.95 y F1-Score: 0.47 mientras que en la otra me da una Exactitud:0.82, Precision:0.84, Sensibilidad:0.95, Especificidad:0.95 y F1-Score: 0.89 .

V. DISCUCION

Bagging se centra en reducir la varianza mediante la combinación de múltiples modelos independientes, mientras que Boosting se centra en reducir el sesgo mediante la construcción iterativa de una secuencia de modelos que corrigen errores anteriores.

VI. CONCLUSIONES

La iniciativa de análisis de datos y aprendizaje automático de Omega Bank representa un enfoque estratégico para abordar los desafíos relacionados con los atrasos en los pagos con tarjetas de crédito. Al identificar las variables relevantes, crear

modelos predictivos avanzados y mejorar la gestión del riesgo crediticio, Omega Bank está posicionando de manera efectiva para mitigar riesgos, mejorar la rentabilidad y fortalecer las relaciones con los cliente

VII. ANEXOS:

```
Matriz de confusión:
[[4449 238]
 [ 835 478]]
Exactitud: 0.82
Precisión: 0.67
Sensibilidad: 0.36
Especificidad: 0.95
F1-Score: 0.47
```

Fig. 1. Prediccion

```
Matriz de confusión:
[[4449 238]
 [ 835 478]]
Exactitud: 0.82
Precisión: 0.84
Sensibilidad: 0.95
Especificidad: 0.95
F1-Score: 0.89
```

Fig. 2. Prediccion 2

VIII. BIBLIOGRAFIA