

# MODELOS Y SIMULACION I – Ingeniería de Sistemas

## PROYECTO AMERICAN EXPRESS PREDICTION - ENTREGA 1

Johan Sebastian Henao Cañas

1000085432

### 1. Descripción del Problema Predictivo

#### *Contexto:*

¿Cómo saben los emisores de tarjetas que devolveremos lo que cobramos? Se trata de un problema complejo con muchas soluciones existentes. En el ámbito de las instituciones financieras, la predicción del incumplimiento crediticio desempeña un papel crítico. Esta predicción permite a las entidades crediticias tomar decisiones informadas sobre la concesión de créditos y gestionar el riesgo de manera eficiente. En particular, la predicción del incumplimiento crediticio es esencial para garantizar la solidez económica, mantener la satisfacción del cliente y reducir pérdidas financieras.

American Express es una empresa de pagos integrada globalmente. Siendo el mayor emisor de tarjetas de pago del mundo, brindan a los clientes acceso a productos, conocimientos y experiencias que enriquecen vidas y fomentan el éxito empresarial.

#### *Objetivo:*

El objetivo de este proyecto es desarrollar un modelo de aprendizaje automático que pueda predecir la probabilidad de que un cliente no pague el saldo de su tarjeta de crédito en el futuro. Esta predicción se basa en el perfil del cliente y considera un período de 18 meses después del último extracto de la tarjeta. Un evento de incumplimiento se define como el no pago del monto adeudado en 120 días después de la última fecha del extracto. La variable objetivo es binaria (1 si hay incumplimiento, 0 si no lo hay).

### 2. Dataset Utilizado

El conjunto de datos utilizado en este proyecto se ha obtenido de una fuente confiable y se proporciona por American Express como parte de la competencia "American Express Default Prediction." en la plataforma Kaggle (El enlace está al final del documento [1]).

El conjunto de datos contiene características de perfil agregadas para cada cliente en cada fecha de estado de cuenta. Estas características son anónimas y normalizadas para proteger la privacidad de los clientes. Las características se clasifican en varias categorías generales, que incluyen variables de morosidad (D\_), gasto (S\_), pago (P\_), equilibrio (B\_), y riesgo (R\_\*). Estas características proporcionan información clave para predecir el incumplimiento crediticio.

Los archivos que encontramos al descargar el conjunto de datos son:

- *train\_data.csv*: datos de entrenamiento (33.16 %) con múltiples fechas de extracto por ID\_de\_cliente.
- *train\_labels.csv*: etiqueta de destino para cada ID\_de\_cliente.
- *test\_data.csv*: datos de prueba (66.83%) correspondientes; su objetivo es predecir la etiqueta de destino para cada ID\_de\_cliente.
- *sample\_submission.csv*: un archivo de envío de muestra en el formato correcto.

El objetivo es la columna “*target*” en el archivo “*train\_labels.csv*”.

### 3. Métricas de Desempeño Requeridas

#### *Métricas de Machine Learning:*

G (Normalized Gini Coefficient): El coeficiente Gini normalizado es una métrica que evalúa la capacidad del modelo para clasificar correctamente a los clientes con incumplimiento crediticio. El Gini Coefficient se basa en la curva ROC (Receiver Operating Characteristic) y mide el área entre la curva ROC del modelo y la línea de referencia (ROC aleatoria). Cuanto mayor sea el valor del coeficiente Gini, mejor será la capacidad de clasificación del modelo.

D (Tasa de Incumplimiento al 4%): Esta métrica evalúa la sensibilidad o capacidad de recall del modelo para detectar incumplimientos crediticios. Representa el porcentaje de etiquetas positivas (incumplimientos) que se capturan dentro de las predicciones clasificadas en el 4% superior. Una alta tasa de incumplimiento al 4% indica una capacidad sólida para detectar incumplimientos.

Ambas métricas G y D se combinan para calcular M, Donde:

$$M = 0.5 * (G + D)$$

M tiene un valor máximo de 1.0. Esto significa que el mejor rendimiento posible es 1.0, lo que indicaría que el modelo es perfecto en la clasificación y en la captura de los incumplimientos.

#### *Métrica de Negocio:*

El impacto real en la gestión del riesgo crediticio y la experiencia del cliente. Puede definirse como:

Incremento en la Eficacia de Gestión del Riesgo: el objetivo principal es maximizar la métrica M, que combina G y D. El valor máximo alcanzable para esta métrica es 1.0, por lo tanto, el rendimiento deseable en producción sería alcanzar la puntuación más cercana posible a 1.0 en esta métrica; esto implica que el modelo debe ser capaz de ordenar adecuadamente las probabilidades de incumplimiento y capturar un alto porcentaje de incumplimientos reales.

### 4. Desempeño Deseable en Producción

La creación de este modelo predictivo de incumplimiento crediticio tiene varias implicaciones en la vida real o en las empresas que opten por hacer uso de las tecnologías para obtener un mayor desempeño. Uno de los beneficios es que, desde la construcción de un modelo preciso, puede ayudar a las empresas a identificar a los clientes con mayor riesgo de incumplimiento, lo que reduce el riesgo

de pérdidas financieras y genera mayor rentabilidad debido a que al tomar decisiones crediticias más precisas, las empresas pueden prestar de manera más eficiente y rentable, maximizando los ingresos y minimizando las pérdidas.

También el cliente se vería beneficiado de cierto modo al utilizar modelos precisos, las empresas pueden tomar decisiones de crédito más rápidas y justas, lo que mejora la experiencia y la satisfacción. Y las empresas al observar los clientes con un bajo riesgo de incumplimiento, estos pueden beneficiarse de tasas de interés más bajas en sus préstamos, lo que puede ahorrarles dinero a lo largo del tiempo y hacer que el crédito sea más asequible.

## 5. Referencias:

- [1] <https://www.kaggle.com/competitions/amex-default-prediction/overview>
- [2] <https://repositorio.uniandes.edu.co/bitstream/handle/1992/43940/u830406.pdf?sequence=1>
- [3] <https://www.kaggle.com/code/batzner/gini-coefficient-an-intuitive-explanation>
- [4] <https://towardsdatascience.com/using-the-gini-coefficient-to-evaluate-the-performance-of-credit-score-models-59fe13ef420>



UNIVERSIDAD  
DE ANTIOQUIA

1 8 0 3