

## **PRIMERA ENTREGA DEL PROYECTO**

Presentado por:

Aura María Molina Amaya  
Valeria González González  
Maryely Isabel Rubio de la Cruz

Presentado a:

Raúl Ramos Pollán



**UNIVERSIDAD DE ANTIOQUIA**  
1803  
**FACULTAD DE INGENIERÍA**

**UNIVERSIDAD DE ANTIOQUIA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL  
INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL PARA LAS CIENCIAS E  
INGENIERÍAS  
2023**

## 1. Planteamiento del problema

Actualmente, muchas las personas que desean adquirir nuevos préstamos no cuentan con un historial crediticio suficiente o simplemente nunca han adquirido un crédito en su vida; por lo cual deben acudir a otros medios para obtener el dinero y es allí donde prestamistas que no son acreditados para esta labor sacan provecho a beneficio propio. Por ello, y teniendo en cuenta una variedad de datos alternativos, incluida la información de telecomunicaciones y transaccional, se predecirá cuán capaz es cada solicitante de pagar un préstamo.

## 2. Dataset

Se utilizará el dataset de una competencia de Kaggle llamada **Home Credit Default Risk** (<https://www.kaggle.com/competitions/home-credit-default-risk>), en la cual, se proporcionan los siguientes archivos.

Como variable objetivo, se tiene TARGET (1 - Cliente con dificultades de pago: tuvo un pago atrasado más de x días en al menos una de las primeras cuotas del préstamo en la muestra, 0 - todos los demás casos).

- **application\_ { train | test } .csv**
  - Esta es la tabla principal, dividida en dos archivos para entrenamiento( con TARGET ) y Prueba ( sin TARGET ).
  - Datos estáticos para todas las aplicaciones. Una fila representa un préstamo en nuestra muestra de datos.
- **bureau.csv**
  - Todos los créditos anteriores del cliente proporcionados por otras instituciones financieras que se informaron a Credit Bureau ( para clientes que tienen un préstamo en la muestra ).
  - Para cada préstamo en nuestra muestra, hay tantas filas como el número de créditos que el cliente tenía en Credit Bureau antes de la fecha de solicitud.
- **bureau\_balance.csv**
  - Saldo mensual de créditos anteriores en Credit Bureau.
  - Esta tabla tiene una fila para cada mes de historial de cada crédito anterior reportado a Credit Bureau –, es decir, la tabla tiene ( # préstamos en muestra \* # de créditos previos relativos \* # de meses donde se tiene algo de historia observable para las filas de créditos anteriores ).
- **POS\_CASH\_balance.csv**
  - Instantáneas del saldo mensual del punto de venta anterior de POS ( ) y préstamos en efectivo que el solicitante tenía con el Crédito de la vivienda.

- Esta tabla tiene una fila para cada mes de historial de cada crédito anterior en Crédito de la vivienda ( crédito al consumo y préstamos en efectivo ) relacionados con préstamos en la muestra –, es decir. la tabla tiene ( # préstamos en muestra \* # de créditos previos relativos \* # de meses en los que tenemos algún historial observable para las filas de créditos anteriores ).
- **credit\_card\_balance.csv**
  - Instantáneas de saldo mensual de tarjetas de crédito anteriores que el solicitante tiene con Home Credit.
  - Esta tabla tiene una fila para cada mes de historial de cada crédito anterior en Crédito de la vivienda ( crédito al consumo y préstamos en efectivo ) relacionados con préstamos en la muestra –, es decir. la tabla tiene ( # préstamos en muestra \* # de tarjetas de crédito anteriores relativas \* # de meses en los que tenemos algún historial observable para las filas de tarjetas de crédito anteriores ).
- **precedente\_application.csv**
  - Todas las solicitudes anteriores para préstamos de crédito a domicilio de clientes que tienen préstamos en nuestra muestra.
  - Hay una fila para cada aplicación anterior relacionada con préstamos en nuestra muestra de datos.
- **installments\_payments.csv**
  - Historial de pagos de los créditos desembolsados previamente en el Crédito de la vivienda relacionado con los préstamos de la muestra.
  - Hay una fila ) por cada pago realizado más b ) una fila cada uno por el pago perdido.
  - Una fila es equivalente a un pago de una cuota o una cuota correspondiente a un pago de un crédito de vivienda anterior relacionado con préstamos de la muestra.
- **HomeCredit\_columns\_description.csv**
  - Este archivo contiene descripciones de las columnas en los diversos archivos de datos.

Para realizar el modelo, se utilizará el archivo con los datos de entrenamiento (application\_train.csv), el cual cuenta con una muestra de 307.511 clientes.

Como lo que se busca predecir es el comportamiento de pago futuro de los clientes a partir de datos de comportamiento crediticio de aplicación, demográficos e históricos; en el dataset, se encuentran columnas con información del cliente como el tipo de préstamo, monto del préstamo, género, edad, si tiene carro y/o casa, número de hijos, nivel educativo, ocupación, ingresos, tipo de ingresos, entre otras, como el cumplimiento de la documentación requerida, etc.

### 3. Métricas

Como métrica se utilizará la asignada por la competencia: el área bajo la curva ROC entre la probabilidad prevista y el objetivo observado.

El AUROC se calcula como el área bajo la curva ROC. Una curva ROC muestra la compensación entre la tasa positiva verdadera (TPR) y la tasa positiva falsa (FPR) en diferentes umbrales de decisión.

Una curva ROC siempre comienza en la esquina inferior izquierda, es decir, el punto (FPR = 0, TPR = 0) que corresponde a un umbral de decisión de 1 (donde cada ejemplo se clasifica como negativo, porque todas las probabilidades pronosticadas son inferiores a 1)

Una curva ROC siempre termina en la esquina superior derecha, es decir, el punto (FPR = 1, TPR = 1) que corresponde a un umbral de decisión de 0 (donde cada ejemplo se clasifica como positivo, porque todas las probabilidades pronosticadas son mayores que 0)

Para calcular el conjunto de pruebas AUROC para una tarea de clasificación binaria, se puede emplear el siguiente código de python:

```
import sklearn.metrics
fpr, tpr, umbrales = sklearn.metrics.roc_curve ( y_true = true_labels,
y_score = pred_probs, pos_label = 1 ) # clase positiva es 1; clase negativa
es 0
auroc = sklearn.metrics.auc ( fpr, tpr )
```

### 4. Desempeño

Puesto que Home Credit se enfoca en préstamos responsables principalmente a personas con poco o ningún historial crediticio. Lo que se espera es que se obtenga un buen modelo de predicción, el cual garantice que los clientes capaces de pagar no sean rechazados y que los préstamos se otorguen con un calendario de capital, vencimiento y reembolso que permita a sus clientes tener éxito.

### 5. Referencias

Kaggle. (s.f.). *Home Credit Default Risk*.

<https://www.kaggle.com/competitions/home-credit-default-risk/overview>

Draeos, R (s.f). *Measuring Performance: AUC (AUROC)*

<https://glassboxmedicine.com/2019/02/23/measuring-performance-auc-auroc/>