

## PROPUESTA DE MONOGRAFÍA

|                     |                      |   |
|---------------------|----------------------|---|
| Título del proyecto | Give Me Some Credit  |   |
| Estudiante 1        |                      |   |
| Nombres completos   | Federico Ocampo      | e-mail: federico.ocampoo@udea.edu.co      |
|                     |                      | GitHub: https://github.com/00Fede         |
| Estudiante 2        |                      |   |
| Nombres completos   | Andres Felipe Orrego | e-mail: andres.orrego2@udea.edu.co        |
|                     |                      | GitHub: https://github.com/felipeorrego97 |

### 1. Descripción del problema

Los bancos juegan un papel muy importante en las economías de mercado. Ellos deciden quien puede ser financiado y bajo qué términos y pueden hacer o declinar decisiones de inversión. Para que los mercados y las sociedades funcionen, los individuos y las compañías necesitan acceso a créditos.

Los algoritmos para calcular el puntaje crediticio, que realizan una estimación de la probabilidad de no pago, son métodos usados por los bancos para determinar si un préstamo debe otorgarse o no.

Un pronóstico de la probabilidad de estrés financiero en los clientes permitiría adelantarse a situaciones donde la cartera vencida aumente, a causa de la mora en los pagos de los créditos, y tomar decisiones al respecto como ofrecer acuerdos de pago, campañas o contacto con los clientes.

### 2. Descripción del dataset

Para este problema se provee una data histórica de cerca de 250.000 deudores. Los datos se distribuyen de la siguiente manera: se cuentan con 150.000 muestras para el entrenamiento del modelo y 101.503 muestras para la evaluación. Se tienen 12 variables predictoras cuantitativas y no hay variables cualitativas. La variable de salida **SeriousDlqin2yrs** es un dato binario que pronostica si una persona tendrá estrés financiero en 2 años o no.

El problema corresponde a la competición subida en la plataforma kaggle Give Me Some Credit, que puede encontrarse en este [link](#). Se entregan 4 archivos: cs-test.csv, cs-training.csv, Data Dictionary.xls y sampleEntry.csv.

A continuación se da una descripción de cada uno de los predictores:

| Variable Name                        | Description  | Type       |
|--------------------------------------|--|------------|
| SeriousDlqin2yrs                     | La persona experimentó 90 días o más de morosidad  | Y/N        |
| RevolvingUtilizationOfUnsecuredLines | Saldo total en tarjetas de crédito y líneas de crédito personales, excepto bienes raíces y sin deuda a plazos, como préstamos para automóviles, dividido por la suma de los límites de crédito | percentage |
| age                                  | Edad en años   | integer    |
| NumberOfTime30-59DaysPastDueNotWorse | Número de veces que el prestatario ha estado atrasado entre 30 y 59 días, pero no peor en los últimos 2 años.  | integer    |
| DebtRatio                            | Pagos mensuales de deuda, pensión alimenticia, costos de vida divididos por el ingreso bruto mensual   | percentage |
| MonthlyIncome                        | Ingresos mensuales   | real       |
| NumberOfOpenCreditLinesAndLoans      | Número de créditos del prestatario, incluyendo tarjetas de crédito y créditos rotativos  | integer    |
| NumberOfTimes90DaysLate              | Número de veces que el prestatario ha estado en mora por 90 días o más.  | integer    |
| NumberRealEstateLoansOrLines         | Número de préstamos hipotecarios y de bienes raíces, incluidas las líneas de crédito con garantía hipotecaria  | integer    |
| NumberOfTime60-89DaysPastDueNotWorse | Número de veces que el prestatario ha estado atrasado entre 60 y 89 días, pero no peor en los últimos 2 años.  | integer    |
| NumberOfDependents                   | Número de dependientes en la familia que se excluyen a sí mismos (cónyuge, hijos, etc.)  | integer    |

### 3. Métricas de desempeño

Como métrica de desempeño se utilizará la métrica AUC-ROC, aprovechando que no es sensible al desequilibrio entre las clases. Para este caso se espera una AUC de 80% para considerar el modelo como aceptable. Con este valor se busca aumentar la sensibilidad, para disminuir los falsos negativos, clientes que van a tener estrés financiero y son clasificados sin estrés financiero en dos años.

Como métrica de negocio se puede determinar la disminución de la cartera vencida al utilizar el modelo como criterio para la aprobación de créditos financieros.

### 4. Criterio de desempeño

Se espera una disminución en la mora crediticia de un 30% después de aplicado el modelo en los clientes con créditos activos de la organización.

### 5. Referencias

Credit Fusion, Will Cukierski. (2011). Give Me Some Credit. Kaggle.  
<https://kaggle.com/competitions/GiveMeSomeCredit>