

# Score de pago y Perfilacion de clientes cooperativa financiera

## PROYECTO INTEGRADOR 2

Juan David Correa Restrepo

Jose Ignacio Escobar Bedoya

Camilo Rivera Bedoya

Daniel Romero Cardona



# AGENDA

01. Entendimiento del negocio

02. Objetivos

03. Entendimiento de los datos

04. Preparacion de los datos

05. Modelado

06. Resultados

# Entendimiento del negocio

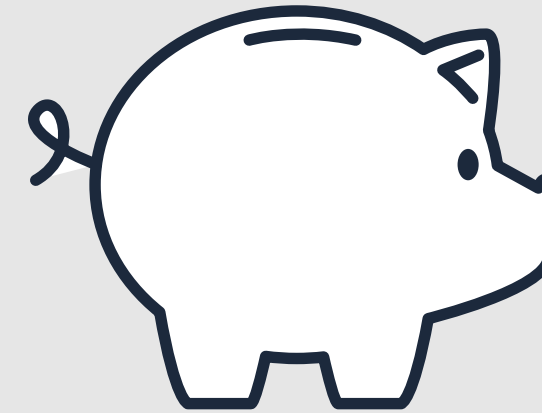
---

## COOPERATIVA

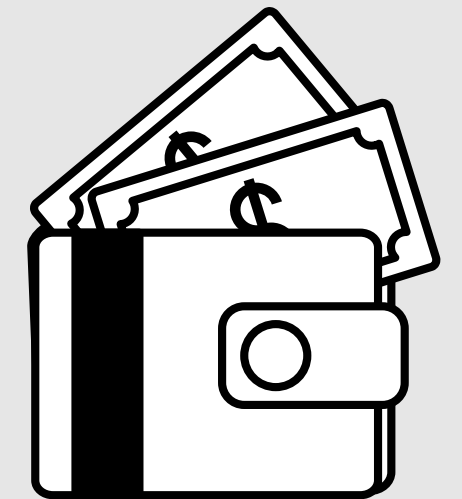


Entidades organizadas para ejercer actividades financieras a través de operaciones activas de crédito

## AHORRO



## CRÉDITO



## ASEGURAMIENTO



# Entendimiento del negocio

---

## Entendimiento del asociado



- Riesgo de cartera
- Venta Cruzada
- Productos a la medida
- Dimensionamiento de productos

**17.324**

Créditos aprobados

**71.256**

Número de asociados

**\$249.849**

Millones de pesos en Créditos

# Objetivo general

---

**Creación de una calificación crediticia basada en la **probabilidad de pago** de los asociados respecto a sus responsabilidades financieras con la cooperativa y que a su vez, sirva como insumo para la **caracterización y perfilación de los diferentes clientes** que se tienen en la entidad.**

# Objetivos Especificos

---

- Utilizar técnicas de machine learning para la clasificación y generación de probabilidad de pago de los asociados.
- Realizar ingeniería de características para la elección y transformación de los datos relevantes para el entendimiento del pago de los asociados.
- Utilización de modelos de aprendizaje no supervisado para entender los diferentes segmentos de clientes que se tienen en la entidad

## Entendimiento de los datos

- Elección de variables
- Extracción

## Preparacion de los datos

- Transformación
- Normalización
- Análisis descriptivo
- Vble. Objetivo

## Modelación

- Cluster puntos atipicos
- Modelos estadisticos
- Validacion cruzada
- Métricas de ajuste

Análisis **segmentación** de asociados



# Entendimiento de los datos

---

## Caracterización cliente

- Ingresos
- Estrato
- Estado civil
- Personas a cargo
- Nivel de escolaridad
- Genero
- Departamento
- Municipio
- Actividad Economica

## Relacionamiento con la cooperativa

- Monto del crédito
- Cuota
- Saldo
- Meses de antigüedad
- Calificación de central de riesgo
- Meses en mora
- Fecha de pago / Fecha pago real
- Valor pagado
- Línea de crédito



# Preparación de los datos

---

**Unificación y filtro de variables**



**Analisis descriptivo**



**Normalización de variables**

## Variable objetivo:

- **Crédito sin mora**
- **Pago completo**
- **Sin atraso en fecha de pago**

# Análisis Descriptivo

## Variables Numéricas

Tendencia central		Posición	
	Resultado		Resultado
Medida		Medida	
Moda	0.00	Mínimo	0.00
Media	1.28	Percentil 1	0.00
Media Armónica	nan	Percentil 5	0.00
Media Geométrica	0.00	Percentil 10	0.00
Media Cuadrática	1.83	Percentil 25	0.00
Media Trunc.(5%)	1.15	Percentil 50	1.00
Media IQ	1.07	Percentil 75	2.00
Media Wins.(5%)	1.24	Percentil 90	3.00
Trimedia	1.00	Percentil 95	4.00
Mediana	1.00	Percentil 99	5.00
Mid Range	5.50	Máximo	11.00
Mid Hinge	1.00		

### Dispersión

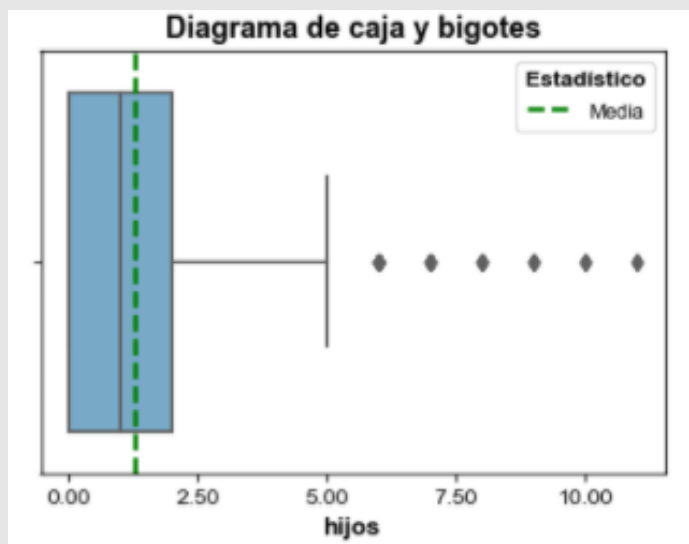
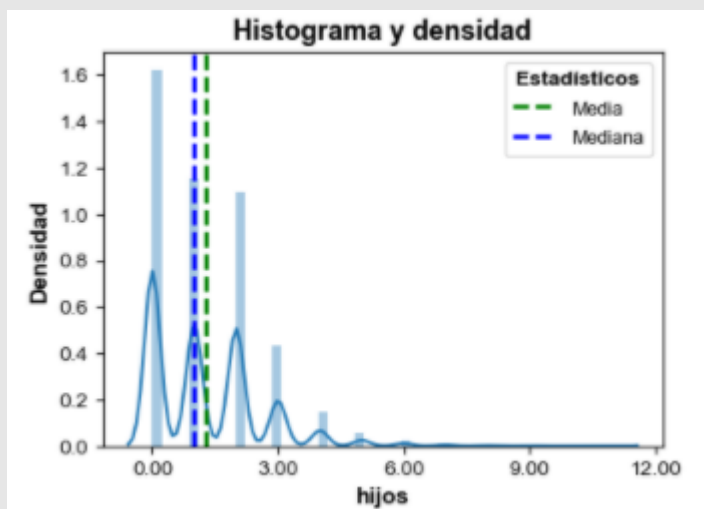
Medida	Resultado
Desv. Est.	1.31
Rango	11.00
Rango IQ	2.00
Dif. Abs. Media	1.05
Dif. Abs. Mediana	1.00
Coef. Var.	1.03
QCD	1.00

### Forma

Medida	Resultado
Asimetría	1.34
Exc.Curtosis	3.31

### Conteo de registros

	Frec.Abs.	Frec.Rel.	Frec.Rel.Acum.
Con dato	20,853	100.00%	100.00%

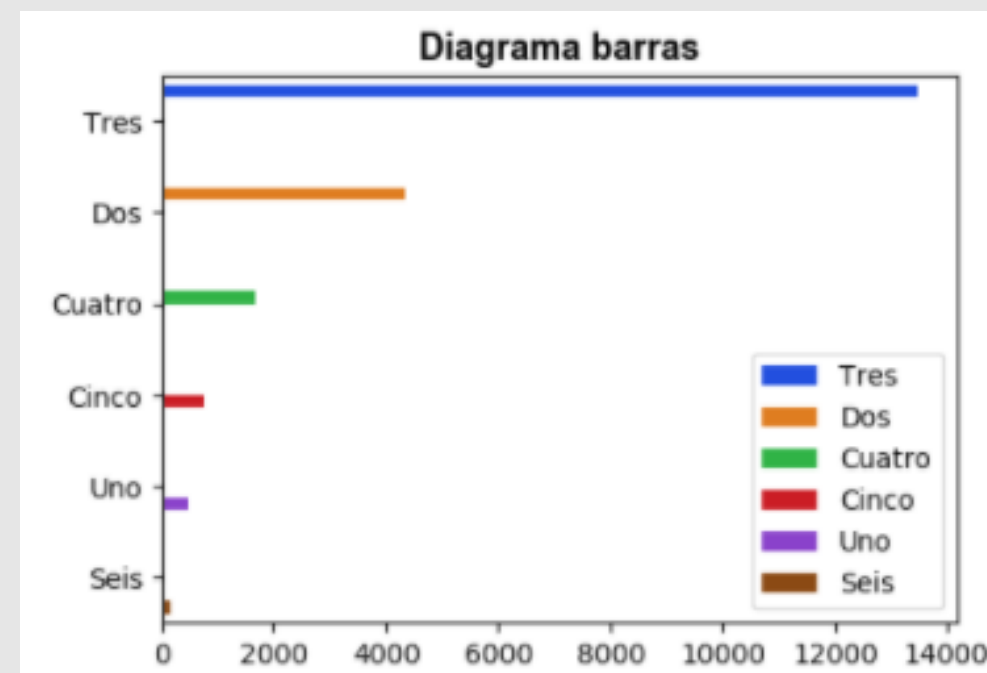


# Análisis Descriptivo

## Variables Categoricas

### Conteo de frecuencias

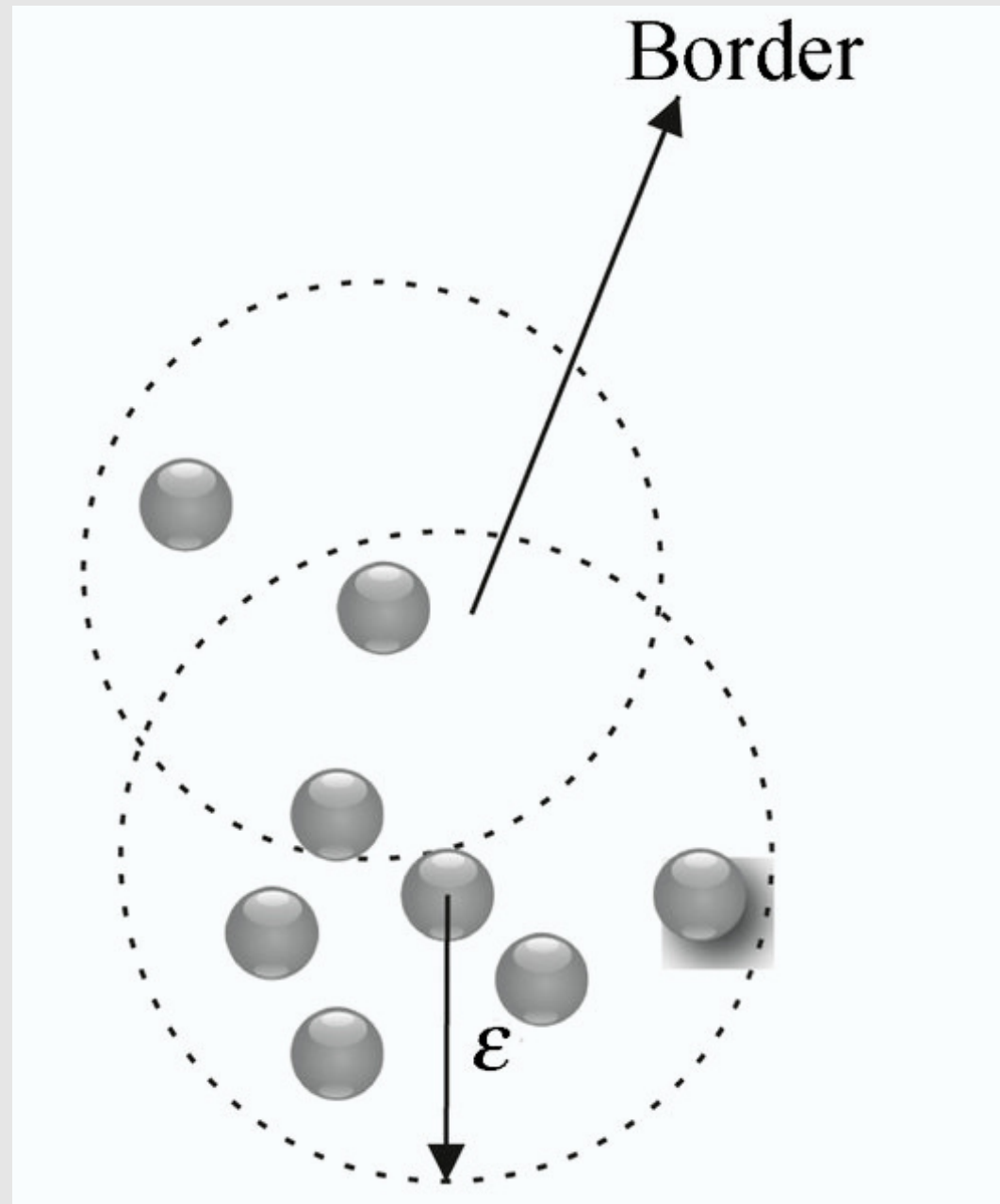
	Frec.Abs.	Frec.Rel.	Frec.Rel.Acum.
Categorías			
Tres	13,495	84.71%	84.71%
Dos	4,322	20.73%	85.44%
Cuatro	1,672	8.02%	93.46%
Cinco	757	3.63%	97.09%
Uno	459	2.20%	99.29%
Seis	148	0.71%	100.00%



## Selección

- Estrato social
- Género
- Estado civil
- Nivel educativo
- Ind\_Personas\_ACargo
- Meses antigüedad
- Ingresos
- Monto crédito

# Modelado



# DBSCAN

## Aprendizaje no supervisado

Identificación de outliers basados en la clusterización de datos multivariados, segmento aquellos comportamientos atípicos que tiene similitud con el resto de los datos presentes.

[1] Ghallab, FahMy & Nasr (2020). Detection outliers on internet of things using big data technology. ScienceDirect Vol. 21 Pg. 131-138

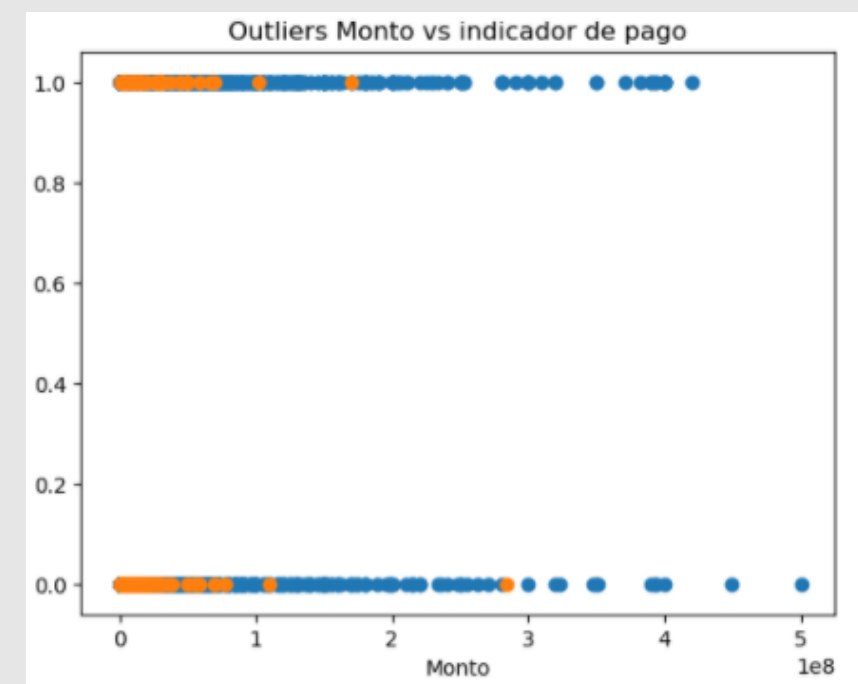
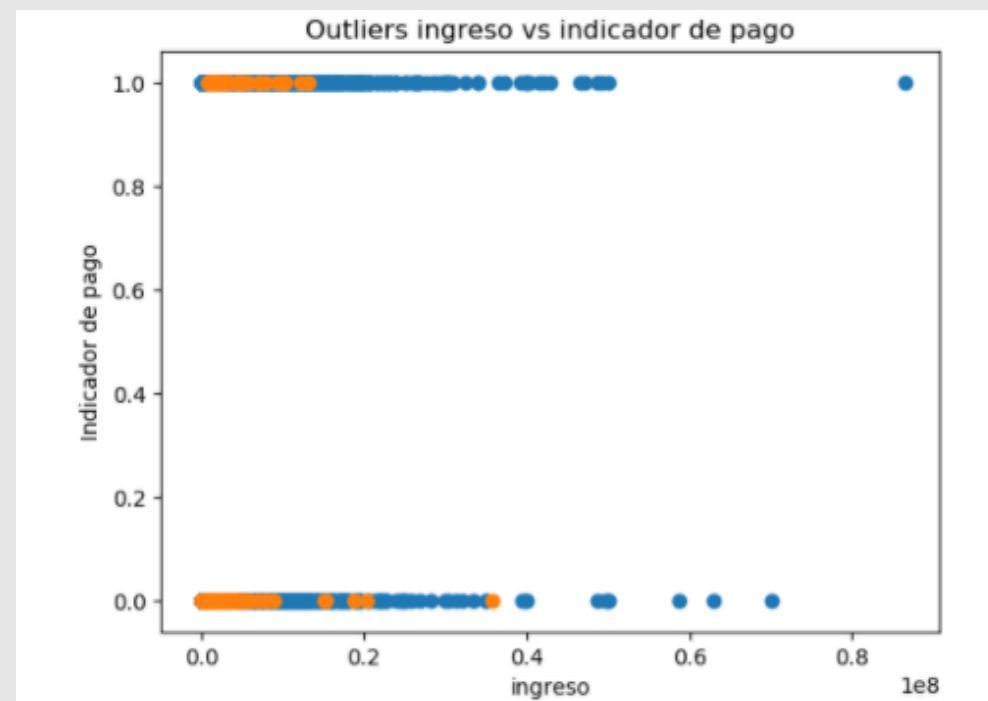
[2] Akbari & Unland (2016). Automated Determination of the input Parameters of DBSCAN Based on Outliers detection. Artificial Intelligence Applications and Innovations Vol. 475 Pg. 280-291

# Modelado

# DBSCAN

Aprendizaje no supervisado

Estrato	Nivel Educativo	Ingreso	Monto
Dos	Primaria	8'878.600	30'000.000
Dos	Primaria	20'300.000	15'000.000
Cinco	Profesional	0	15'500.000



# Modelado

## Entrenamiento

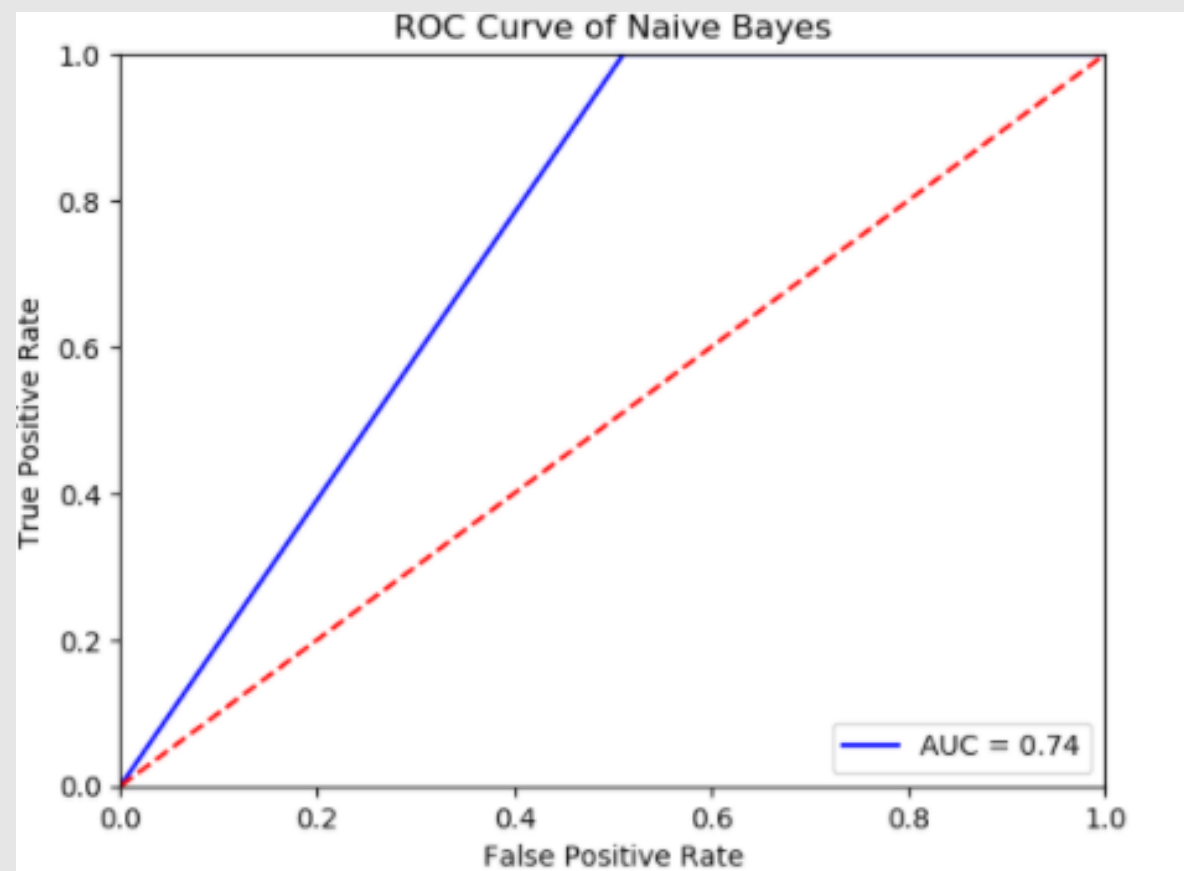
Cross Validation: 10 Kfolds

## Prueba

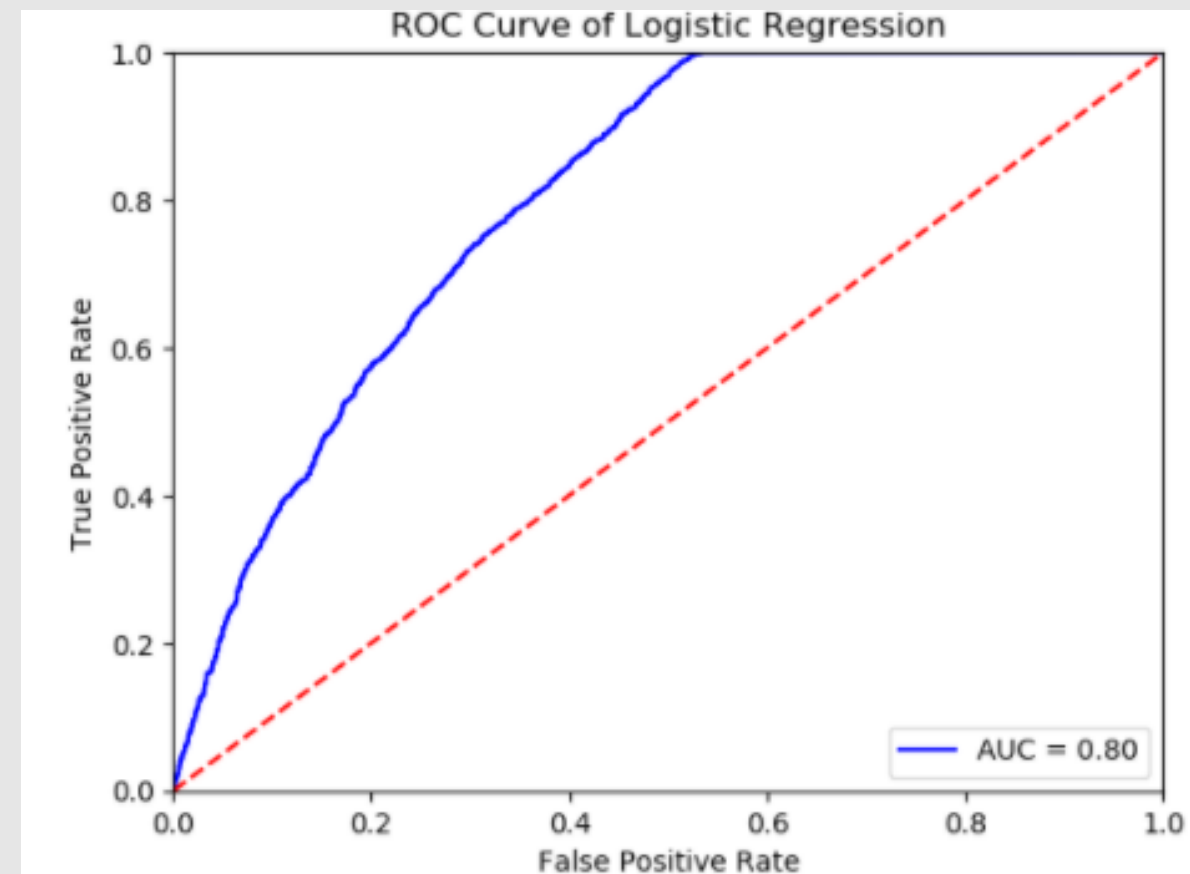
Modelo	Accuracy	Recall	F1_Score	Accuracy	Recall	F1_Score
Regresión logística	0.75	0.99	0.82	0.77	0.99	0.83
Naive Bayes	0.76	1.0	0.82	0.78	1.0	0.84
D. Tree	0.71	0.73	0.74	0.71	0.75	0.75
SGD	0.74	0.95	0.80	0.74	0.87	0.79
KNN	0.62	0.72	0.68	0.63	0.72	0.69
SVM	0.68	0.99	0.77	0.70	0.99	0.79

# Modelado

**ROC- AUC Naive Bayes**



**ROC- AUC Regresion Logistica**



$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}$$

**SCORE DE PAGO: [0-1]**



# Segmentacion clientes

---

- **K-MEANS (Distancias)**
- **Cluster Espectral (Matriz afinidad)**
- **Ward (Jerarquico)**

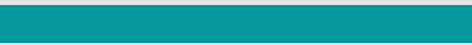
## Selección

- **Estrato social**
- Género
- Estado civil
- **Nivel educativo**
- Ind\_Personas\_ACargo
- Meses antigüedad
- **Ingresos**
- Monto crédito



Las variables socio-demograficas que se recolectan de los afiliados no son suficientes para hacer un analisis **concluyente** discriminante de los perfiles de los afiliados en cuanto a su comportamiento de pago. **Clases dominante homogeneas en todos los cluster**

# Perfilación clientes



## Estado Civil

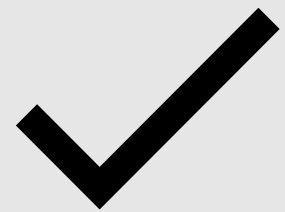
P(Pago)	Casado	Separado	Soltero	U. Libre	Viudo	N.A.
< 0.5	4477	830	4331	1528	558	0
> 0.5	3243	705	3493	1311	375	2

## Estrato Socioeconomico

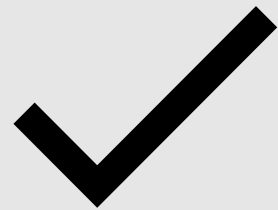
P(Pago)	Uno	Dos	Tres	Cuatro	Cinco	Seis
< 0.5	256	2321	7766	901	406	74
> 0.5	203	2001	5729	771	351	74

# Resultados

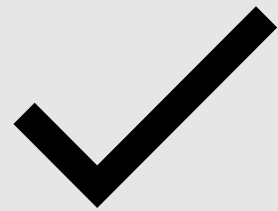
---



**Método robusto de clientes con comportamiento atípico. (A. No supervisado)**



**Score - Calificación comportamiento de pago. (A. Supervisado)**



**Análisis descriptivo \*\* de perfil de cliente respecto a su comportamiento de pago**

GRACIAS !

