

## Introduccion y Problematica

Arca Continental, una de las principales embotelladoras de Latinoamérica, busca innovar en mercadotecnia mediante técnicas avanzadas de análisis de datos.

Uno de los principales desafíos de la empresa es identificar a los clientes con mayor probabilidad de adoptar un nuevo producto, optimizando así sus estrategias de lanzamiento y aumentando su impacto en el mercado.

## Objetivos

Analizar las características más relevantes de los clientes para proponer una segmentación adecuada

Comparar diferentes modelos predictivos para identificar el más adecuado para la selección de clientes.

Definir métricas y técnicas de evaluación adecuadas para determinar la funcionalidad de los modelos.

## Metodología

### Procesamiento de Datos

Los clientes de Arca incluyen tiendas de diversos tamaños, y las bases de datos proporcionadas ofrecen un panorama detallado del entorno de cada cliente, los productos que adquieren y sus características. Tras analizar la información, se eliminaron redundancias, se simplificó la complejidad mediante la transformación de variables a representaciones más significativas y se generaron nuevas características clave.

Análisis de distribuciones

Análisis de dispersión de variables

Clusterización de clientes en base a sus características

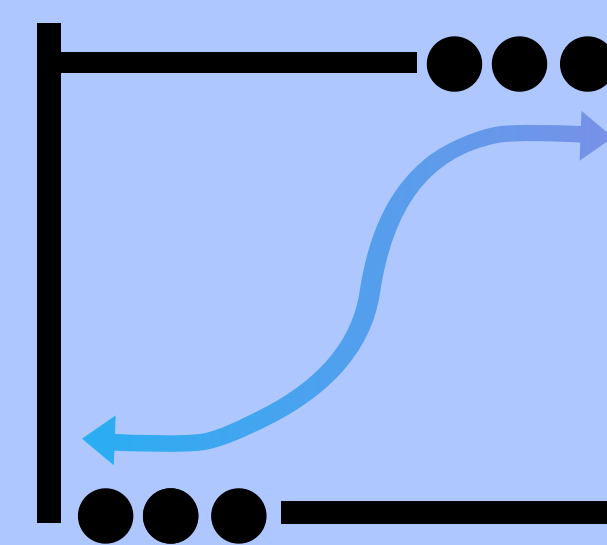
Tratamiento de datos faltantes

### Modelos

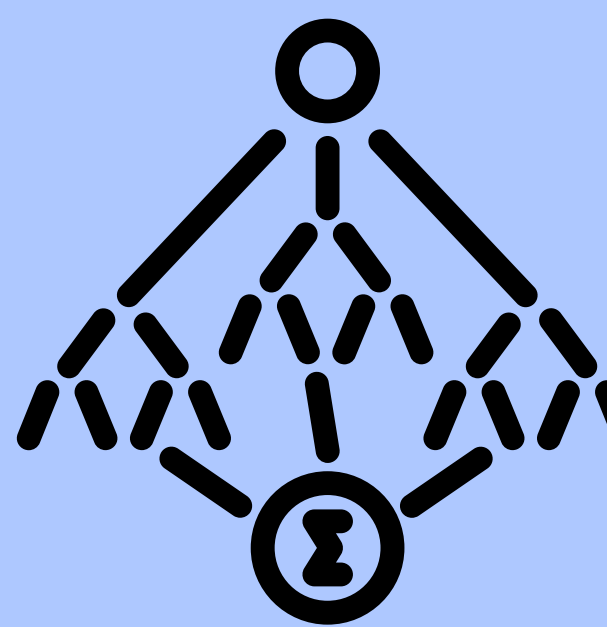
Se preparan los datos para ser ingresados a un modelo clasificador, para lo cual se realizaron las siguientes actividades:

- Transformación de variables categóricas (One-Hot Encoding)
- Balanceo de clases utilizando SMOTE (Synthetic Minority Oversampling Technique)
- Se escalan las variables utilizando media y desviación estándar.

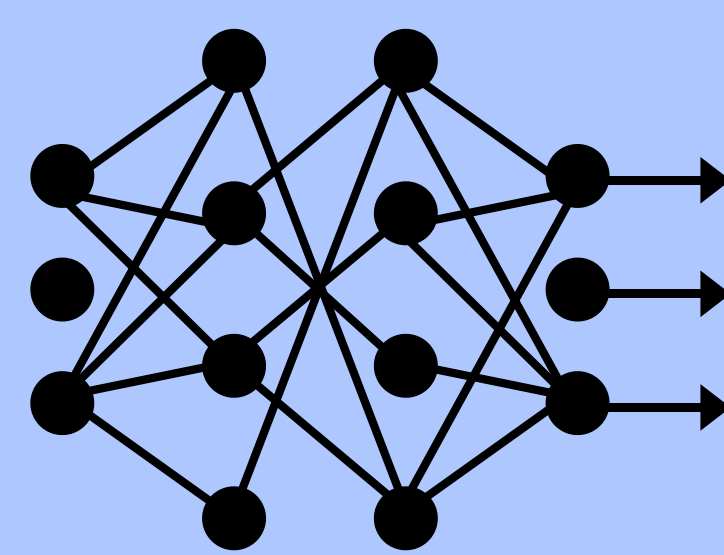
Regresión Logística



Random Forest y XGBoost



Neural Network



### Metricas de Evaluacion

Accuracy

ROC AUC

Recall

Matrices de Confusión

## Resultados

ID Cliente	Probabilidad
1	0.91
2	0.85
3	0.81
4	0.78
5	0.65

### Evaluación

Se consideraron 2 productos del dataset de prueba y se generaron las probabilidades para los clientes, se tomó el top (5, 10, 25, 50, 100) y se calcula la proporción de clientes que compraron en al menos 2 ocasiones.

#### Evaluación Deseada

Top Clientes	Proporción de compras
5	0.80
10	0.60
25	0.50
50	0.30
100	0.25

#### Evaluación No Deseada

Top Clientes	Proporción de compras
5	0.00
10	0.10
25	0.04
50	0.06
100	0.05