

# Sprint 3: Perfilado Inteligente de Clientes para Maximizar la Suscripción de Depósitos a Plazo



## Introducción

Este proyecto tiene como objetivo crear un sistema de **perfilado inteligente de clientes** basado en las características demográficas, financieras y de interacción para predecir qué clientes tienen más probabilidades de suscribir un depósito a plazo fijo. En lugar de solo predecir "sí" o "no", el proyecto buscará **segmentar a los clientes en diferentes grupos de susceptibilidad** y personalizar las estrategias de marketing para maximizar el éxito de las campañas.

La idea clave aquí es usar Machine Learning para generar perfiles detallados que permitan al banco no solo predecir la suscripción, sino **entender por qué ciertos grupos de clientes tienden a aceptar o rechazar** las ofertas de depósitos a plazo.

## Objetivos del Proyecto

1. **¿Cuáles son los objetivos del negocio?**
  - Identificar los perfiles de clientes más propensos a suscribir un depósito a plazo y desarrollar estrategias personalizadas para cada segmento, lo que mejorará la efectividad de las campañas de marketing directo.
2. **¿Qué decisiones o procesos específicos desea mejorar o automatizar con ML?**
  - Automatizar el perfilado de clientes y la personalización de campañas de marketing para que el equipo de ventas pueda adaptar las ofertas a las características específicas de cada segmento de clientes.
3. **¿Se podría resolver el problema de manera no automatizada?**
  - Aunque la predicción del gasto anual de cada cliente se puede realizar utilizando hojas de cálculo o software estadístico, el uso de machine learning permitirá automatizar este proceso mediante entrenamientos programados que incorporen rápidamente las transacciones diarias sin necesidad de intervención manual.

## Metodología Propuesta

El enfoque propuesto combina dos técnicas principales: **Clustering** y **Clasificación** para identificar grupos de clientes con características similares y predecir la probabilidad de que suscriban un depósito a plazo fijo dentro de cada grupo.

**1.Clustering: Identificación de Grupos de Clientes:** Utilizaremos técnicas de clustering no supervisado para identificar grupos de clientes que comparten características demográficas y financieras, como edad, saldo promedio, historial de préstamos y número de interacciones con el banco.

Los algoritmos seleccionados pueden ser:

- a. **K-means:** Agrupa a los clientes minimizando la distancia dentro de cada grupo y maximizando la distancia entre grupos. Ideal para segmentaciones claras y homogéneas.
- b. **DBSCAN:** Se utilizará si los datos presentan patrones complejos o ruido (clientes que no encajan fácilmente en grupos), ya que este algoritmo agrupa por densidad.

El resultado será una serie de **segmentos** de clientes, por ejemplo:

- **Grupo 1:** Clientes jóvenes con saldo promedio bajo pero alta interacción.
- **Grupo 2:** Clientes mayores con saldo alto y poca interacción.

## 2. Clasificación: Predicción de Suscripción a Depósitos a Plazo Fijo

Una vez identificados los grupos, aplicaremos un modelo de clasificación supervisado para predecir la probabilidad de que los clientes de cada grupo suscriban un depósito a plazo. Los algoritmos a emplear serán:

- XGBoost:** Un algoritmo de boosting, altamente eficiente para capturar relaciones complejas entre variables.
- Random Forest:** Un modelo basado en múltiples árboles de decisión, robusto y eficiente, ideal para conjuntos de datos con múltiples variables.

Las características utilizadas para la predicción incluirán el historial de interacciones con campañas, número de contactos previos, y duración de las llamadas, entre otros factores.

Para evaluar la efectividad del modelo de clasificación, se utilizarán las siguientes métricas:

- **Precisión (Accuracy):** Porcentaje de predicciones correctas sobre el total de predicciones. Útil para una visión general, aunque puede no ser suficiente en casos de desequilibrio de clases.
- **F1-Score:** Combina precisión y recall en una sola métrica, útil cuando hay un desequilibrio entre las clases (pocos "sí" en comparación con "no"). Un buen F1-Score refleja un equilibrio entre la tasa de verdaderos positivos y falsos negativos.
- **ROC-AUC (Área Bajo la Curva ROC):** Mide la capacidad del modelo para distinguir entre clientes que suscriben y los que no. Un AUC cercano a 1 indica excelente rendimiento.
- **Precisión por grupo:** Mide la precisión del modelo dentro de cada segmento de clientes identificado en el clustering, asegurando que la clasificación sea efectiva para todos los grupos.

## Datos Disponibles

El conjunto de datos incluye 45,211 observaciones con 16 variables, que abarcan aspectos como:

- **Características demográficas:** Edad, estado civil, nivel educativo.
- **Características financieras:** Saldo promedio anual, historial de préstamos.
- **Interacción con campañas:** Número de contactos previos, duración de las llamadas, y resultado de campañas anteriores.

Estas variables permitirán crear perfiles detallados y aplicar los modelos de clasificación y segmentación de manera efectiva.

## Métrica de Éxito del proyecto

**Tasa de conversión por perfil:** El porcentaje de clientes que aceptan la oferta dentro de cada segmento, en comparación con los resultados generales. Esta métrica mide si el perfilado y la personalización de estrategias aumentan efectivamente las tasas de conversión de depósitos a plazo fijo.

## Responsabilidades Éticas y Sociales

Como en todo proyecto que involucra datos personales, es necesario tener en cuenta:

- **Privacidad:** Asegurar que se respeten todas las normativas de protección de datos y garantizar que los datos de los clientes estén protegidos y utilizados con consentimiento explícito.
- **Transparencia:** Los clientes deben saber que sus datos se están utilizando para mejorar su experiencia, evitando cualquier sensación de manipulación.
- **Evitar sesgos:** Es crucial revisar los modelos para asegurarse de que no haya sesgos en las predicciones que puedan discriminar a ciertos grupos de clientes.