

Sprint 7.2: 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:

La **Regresión Logística** se ha elegido como el modelo principal para resolver este problema de clasificación binaria. Este algoritmo permite calcular la probabilidad de que un cliente suscriba un depósito, lo cual es fundamental para personalizar estrategias de marketing y mejorar la tasa de suscripción. Además, la regresión logística proporciona coeficientes que permiten identificar la importancia de cada variable, ayudando a entender los factores más influyentes en la decisión de suscripción de los clientes.

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:

- **Exactitud (Accuracy):** La exactitud mide el porcentaje de predicciones correctas en el conjunto de datos. Si bien es útil como referencia general, en este contexto bancario es importante combinarla con otras métricas, especialmente en situaciones de desbalanceo de clases, donde una mayoría de clientes podría optar por no suscribir, haciendo que un modelo sesgado hacia el "no" presente alta exactitud pero bajo rendimiento en el negocio.
- **F1-Score:** Dado que este proyecto busca captar clientes que probablemente suscriban, es importante que el modelo no solo sea preciso, sino también que minimice los falsos negativos (clientes interesados clasificados incorrectamente como no interesados). El F1-Score, al

equilibrar precisión y recall, proporciona una visión más justa del modelo cuando existen clases desbalanceadas, permitiendo identificar mejor a los clientes potenciales para maximizar la tasa de suscripción.

- **ROC-AUC (Área Bajo la Curva ROC):** La métrica AUC-ROC ayuda a evaluar la capacidad del modelo para discriminar entre clientes interesados y no interesados. Esta métrica es ideal en este caso, ya que un alto AUC sugiere que el modelo es eficaz en priorizar clientes interesados en suscribir un depósito. Un valor de AUC más cercano a 1 implica que el modelo es robusto para distinguir entre ambas clases, lo cual es clave para diseñar campañas efectivas.

Además de las métricas de evaluación principales, como el ROC-AUC, F1-Score y precisión, se incluirá **una matriz de confusión** como herramienta complementaria para evaluar el rendimiento del modelo de regresión logística. La matriz de confusión permite desglosar las predicciones del modelo en verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.

Esto es particularmente relevante en el contexto del negocio bancario, donde es crítico minimizar los falsos negativos (clientes que se predijo que no suscribirían pero sí están interesados), ya que representan oportunidades de negocio perdidas. Además, controlar los falsos positivos ayuda a reducir costos en esfuerzos de marketing mal dirigidos. La matriz de confusión permitirá identificar el tipo de errores específicos del modelo y ajustar hiperparámetros si es necesario para lograr un balance óptimo entre precisión y recall.

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.