

PROYECTO DE ML PARA IDENTIFICAR LOS PERFILES DE CLIENTES QUE SE SUSCRIBEN A LOS DEPÓSITOS A PLAZO

INTRODUCCIÓN

Una institución bancaria portuguesa realiza diversas campañas con la finalidad de que los clientes se suscriban a depósitos a plazo. Durante estas campañas, los clientes reciben llamadas telefónicas y deciden si desean suscribirse o no a la promoción ofrecida.

Con el fin de diseñar campañas de marketing más focalizadas, eficientes y menos costosas, es fundamental comprender mejor el perfil de los clientes que finalmente contratan el depósito. Para ello, se pueden aprovechar los detalles demográficos y financieros registrados en la base de datos del banco de cada uno de los clientes contactados.

OBJETIVOS DEL PROYECTO

1. *¿Cuáles son los objetivos del negocio?*

Aumentar la eficiencia de las campañas de marketing de depósitos a plazo, dirigiendo los esfuerzos hacia los clientes con mayor probabilidad de suscripción. Esto, a su vez, reducirá el costo de las campañas.

2. *¿Qué decisiones o procesos específicos desean mejorar o automatizar con ML?*

Automatizar la identificación de perfiles de clientes que son más propensos a suscribirse a los depósitos a plazo y ver qué características (en términos demográficos y financieros) influyen más.

El modelo ayudará al equipo de marketing a identificar automáticamente los clientes que deberían priorizarse en las campañas.

3. *¿Se podría resolver el problema de manera no automatizada?*

Sí, hoy en día se realizan segmentaciones de clientes con diferentes programas, pero puede derivar en un proceso demasiado manual o que no detecte patrones complejos entre las diversas variables que influyen en las decisiones de los clientes.

METODOLOGÍA PROPUESTA

4. *¿Cuál es el algoritmo de Machine Learning más adecuado para resolver este problema?*

Utilizaría algoritmos de clasificación que permiten identificar diferentes grupos de clientes.

Para un primer acercamiento al proyecto, podría utilizar un modelo de *clustering* para identificar subgrupos de clientes con distintas probabilidades de suscribirse a los depósitos.

A continuación, podría utilizar una Regresión Logística, un modelo sencillo y fácil de interpretar, que puede darme pistas sobre qué variables son más influyentes.

Para complementar y mejorar la precisión de identificación de los perfiles, utilizaría el modelo Random Forest, que se conoce por ser más robusto y capaz de manejar datos con características complejas y no lineales. Este modelo, además, proporciona la métrica *feature importance*, que sirve para ver la importancia que tienen las variables demográficas y financieras al final del entrenamiento.

Para evaluar el rendimiento del modelo, definiría la matriz de confusión y calcularía las métricas más importantes en este tipo de modelos, como:

- **Exactitud:** permite calcular la proporción de predicciones correctas respecto al total; las clases están bastante bien repartidas (47.4% contrataron el depósito, 52.6% no lo contrataron), así que esta métrica puede proporcionar información valiosa.
- **Precisión:** permite medir la proporción de verdaderos positivos frente a la totalidad de predicciones positivas, es decir, indica el porcentaje de predicciones positivas que fueron correctas.
- **Sensibilidad (*recall*):** permite medir la proporción de verdaderos positivos en relación con el total de ejemplos positivos presentes en los datos de prueba. En otras palabras, mide la capacidad del modelo para identificar correctamente las muestras positivas. Esto es especialmente interesante porque necesito que los clientes que se suscribieron sean identificados correctamente, y así dirigir los esfuerzos de campañas de marketing a estos perfiles.
- **F1 score:** combina precisión y sensibilidad, proporcionando una medida general del rendimiento del modelo.

Además de interpretar estas métricas, sería interesante asegurarse que los modelos no están sub o sobreajustados e incluso realizar una validación cruzada.

DATOS DISPONIBLES

5. ¿Qué datos están disponibles para abordar este problema?

El banco cuenta con la información demográfica (edad, empleo, estado civil y nivel educativo) y financiera (balance en la cuenta, estado de incumplimiento de crédito, si tienen préstamo hipotecario y/o personal) de los clientes que fueron contactados durante las campañas.

MÉTRICA DE ÉXITO

6. *¿Cuál es la métrica de éxito para este proyecto?*

El éxito de este proyecto se podría evaluar mediante el incremento de la tasa de conversión de la campaña dirigida a los perfiles identificados como los más propensos a suscribirse a la campaña. También se podría medir la reducción en el número de clientes contactados que no se suscriben.

RESPONSABILIDADES ÉTICAS Y SOCIALES

7. *¿Qué responsabilidades éticas y sociales es importante tener en cuenta?*

La responsabilidad principal es asegurarse que los datos personales de los clientes no están sesgados ni discriminan a ciertos grupos demográficos, es decir, que se utiliza información variada que no comprometa a determinados colectivos más vulnerables, ya sea en términos demográficos como financieros.

Además, al tratar con datos personales, es importante que se sigan políticas de privacidad adecuadas para no provocar *leakage* de datos. Es esencial cumplir con el RGPD y otras normativas de protección de datos, garantizando así que los datos de los clientes estén protegidos y su consentimiento para el uso de la información sea respetado. Relacionado con esto, hay que asegurar la transparencia del proceso de uso de datos, mediante políticas de uso de datos deben explicadas de manera clara y comprensible.