

## Nombre del proyecto: Predicción de suscripciones a depósitos a término.

### Introducción

En este modelo se busca predecir si una persona es susceptible a querer contratar un depósito a término o no. La optimización de este servicio es importante ya nos permitiría asignar los recursos justos y necesarios dentro del banco, de igual forma se podría aumentar la imagen del banco, ya que se buscaría contactar exclusivamente a las personas que cumplen con el perfil y por lo tanto considerarían este producto como algo beneficioso en sus vidas y al mismo tiempo se evitaría comunicaciones innecesarias con los clientes que no estarían interesados y por lo tanto podrían considerar las llamadas como molestas.

### Objetivos del Proyecto

1. ¿Cuáles son los objetivos del negocio?  
Predecir un cliente estaría interesado en contratar un depósito a término o no y a partir de allí enfocarse en ese tipo de clientes para aumentar la tasa de suscripción, sería interesante saber que tipo de productos podrían interesarle a los otros tipos de clientes pero esto es materia para otro análisis.
2. ¿Qué decisiones o procesos específicos se desean mejorar o automatizar con ML?  
Se podría mejorar y automatizar la identificación de perfiles de usuarios con alta probabilidad de contratación del producto.
3. ¿Se podría resolver el problema de manera no automatizada?  
Si claro, así se está haciendo hasta el momento, el problema con hacerlo de esta manera es que se puede molestar a clientes que no van a contratar el producto, se contrata a personal extra que puede tener poca eficiencia y por lo tanto se crea mayor frustración en el equipo de ventas, la selección de posibles clientes puede estar muy asociado al conocimiento personal o a la utilización de carteras de clientes, lo cual puede llevar a prácticas poco éticas.

### Metodología Propuesta

Como tengo 17 columnas y muchas filas inicialmente vería si es mejor usar árboles de decisión o bosques aleatorios de clasificación (usando poutcome como la variable categórica a predecir). Para saber si el modelo generado es correcto usaría la métrica F1 score, ya que el objetivo es identificar a los clientes interesados, pero como métricas de soporte buscaría saber como están la precisión (los verdaderos positivos vs el total de positivos predichos) y la sensibilidad (los verdaderos positivos vs los positivos reales), también podría usar una matriz de confusión para entender como el algoritmo generado es capaz de acertar las predicciones reales.

Creo que empezaría con el árbol de decisión y luego haría un modelo de bosques y chequearía si hay problemas importantes de errores y sobreajustes. Si los errores y el sobreajuste no son importantes usaría un modelo de árbol de decisión para usar menos recursos.

## Datos Disponibles

Cuento con un dataset que tiene 17 columnas ('age', 'job', 'marital', 'education', 'default', 'balance', 'housing', 'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pdays', 'previous', 'poutcome', 'deposit') y más de 11 mil filas.

## Métrica de éxito del proyecto

La primera métrica que usaría es porcentaje de clientes que contratan el servicio por cantidad de clientes contactados, luego esto puede traducirse a otro tipo de indicadores de acuerdo al área, por ejemplo al equipo de ventas le interesaría saber el ratio de contrataciones por llamada al contactar usuarios que recomienda el algoritmo vs usuarios que son contactados de acuerdo a la metodología anterior (para evaluar la efectividad del proyecto), al equipo de finanzas le interesaría saber como el modelo se refleja en un aumento de ventas diarias, a los responsables de operaciones del call center le interesaría saber el aumento de la eficiencia por turno o trabajador, a la gente de calidad le gusta saber como la realización de llamadas a clientes que están interesados en el producto aumenta la posibilidad de recomendar el banco o el producto a otras personas, y así esta métrica podría traducirse de diferentes maneras en las diferentes áreas del banco.

## Responsabilidades Éticas y Sociales

Hay que tomar en cuenta que las bases de datos desde donde se originan los perfiles de clientes, se generan a partir de clientes que han aceptado que se use su información para este tipo de estudios y que la información que se usa es exclusivamente la que ellos han aceptado que se use. También hay que tomar en cuenta de donde se saca la información de los posibles clientes, hay que asegurarse que estas fuentes de información son confiables y además obtuvieron la información de manera legal y contando con la autorización previa de las personas.

Hay que asegurarse que ni la información ni el modelo van a sufrir filtración a otras organizaciones, que pudieran poner en riesgo la privacidad de los clientes.

Antes de que ocurra cualquier filtración o hecho que ponga en riesgo la privacidad de la información, la organización debería contar con políticas y procesos claros que permitan establecer los límites y responsabilidades de todo el personal involucrado, y el código ético con el que se tratará toda la información que se genera a lo largo del ciclo de vida del modelo.

El link a mi github es

<https://github.com/isashi04/MachineLearningT3.2/tree/main>