

**Nombre del proyecto:** Proyecto de ML para incrementar la efectividad de las campañas de marketing bancarias para la colocación del producto depósito a plazo.

## Introducción

Los datos están relacionados con campañas de marketing directo (llamadas telefónicas) de una institución bancaria portuguesa. El objetivo de la clasificación es predecir si el cliente suscribirá un depósito a plazo (variable y).

## Objetivos del Proyecto

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

Determinar la probabilidad de comprar del producto depósito a plazo por parte de un cliente dadas sus características y de esta forma enfocar los esfuerzos del equipo comercial para incrementar la tasa de éxito de las campañas de marketing.

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

A partir de los datos generados diariamente por los equipos comerciales mediante llamadas telefónicas, se pretende automatizar el procesamiento de esta información para generar conclusiones sobre la tasa de éxito de una venta y tomar decisiones estratégicas para el cumplimiento de los objetivos comerciales, al igual que personalizar la propuesta de valor del banco ofreciendo un portafolio de productos financieros que incluya el depósito a plazo a los clientes que cuentan con mayor probabilidad de requerirlo.

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

Dada la naturaleza de generación de datos diarios de una gran cantidad de agentes de ventas y clientes, las betas que ponderan el impacto de las variables independientes, constantemente se modifican por los gustos y preferencias de los consumidores, la situación económica que este atravesando la zona de influencia entre otros aspectos, este es un problema que requiere necesariamente una solución automatizada que sea capaz de capturar esos cambios a través del tiempo y brinde una solución alineada con la velocidad en la que se desarrolla específicamente el sector bancario.

## Metodología Propuesta

### Breve descripción del tipo y algoritmo de ML a utilizar.

El algoritmo más adecuado para resolver este problema es el **Regresor de Bosques Aleatorios (Random Forest Regressor)**. Este algoritmo es eficaz para predecir valores numéricos, como el promedio en cuenta del último año, la duración de las llamadas y los intentos de contacto realizados, porque maneja bien datos complejos y relaciones no lineales. Además, es robusto frente a valores atípicos y reduce el riesgo de sobreajuste (overfitting) al promediar los resultados de múltiples árboles de decisión y puede manejar variables categóricas como es el caso del job, housing, education, etc.

### Las métricas de evaluación que se utilizarán son:

**Error Cuadrático Medio (MSE | Mean squared error):** Para medir la diferencia entre los valores predichos y los reales.

**R<sup>2</sup> (Coeficiente de Determinación):** Para evaluar qué tan bien el modelo explica la variabilidad del gasto anual.

### Teniendo en cuenta que las features disponibles incluyen:

age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, poutcome, deposit.

La justificación de que el Regresor de Bosques Aleatorios (Random Forest Regressor) es una excelente elección recae en varias razones:

- **Capacidad de manejar tanto variables numéricas como categóricas:** Random Forest puede trabajar fácilmente con datos numéricos como el tiempo de interacción y el promedio de depósitos por año, así como con variables como el nivel educativo o trabajo. No requiere un extenso preprocesamiento o transformación de los datos categóricos, ya que los árboles de decisión dentro del bosque los manejan de manera natural.
- **Reducción del riesgo de sobreajuste (overfitting):** Dado que el modelo utiliza múltiples árboles de decisión (cada uno entrenado con diferentes subconjuntos de los datos), promedia los resultados, lo que reduce el riesgo de que un solo árbol aprenda demasiado los detalles específicos del conjunto de datos (sobreajuste) y pierda capacidad de generalización.
- **Tratamiento de relaciones no lineales:** Las interacciones entre el tiempo llamada y el nivel educativo podrían tener efectos complejos en la probabilidad de que un cliente tome un depósito a término. Random Forest maneja bien estas relaciones no lineales, que serían difíciles de capturar con modelos más simples como la regresión lineal.
- **Robustez frente a datos faltantes o ruidosos:** Si hay datos incompletos o valores atípicos en las interacciones de los clientes o el historial de compras, Random Forest puede ser más tolerante y aun así ofrecer buenas predicciones sin necesidad de eliminar o imputar datos de manera intensiva.
- **Importancia de las características:** Random Forest ofrece una forma de medir la importancia de las características, lo que ayuda a identificar cuáles de las variables que tienen un mayor impacto en la probabilidad de éxito. Esta capacidad de análisis es muy útil para la empresa, ya que permite enfocar los esfuerzos en clientes con características específicas, y crear campañas a medida.

### Métrica de éxito del proyecto

Incrementar el porcentaje de efectividad de una campaña de marketing en un 30%: Medida a través de la cantidad de depósitos a termino colocados en un periodo de tiempo determinado.

### Responsabilidades Éticas y Sociales

**Privacidad y protección de datos:** El modelo utilizará datos personales e identificativos de los clientes.

**Transparencia en las decisiones automatizadas:** Es importante que las predicciones los clientes no sean utilizadas de manera discriminatoria o para limitar las oportunidades de algunos clientes. El modelo debe ser transparente y auditable, de forma que se pueda explicar cómo se toman las decisiones y evitar sesgos que perjudiquen a ciertos grupos de clientes.

**Evitar el sesgo algorítmico:** Asegurarse de que el modelo no favorezca a un tipo específico de cliente. El sesgo en el modelo podría generar desigualdad en el trato y experiencia de compra de los clientes.

**Impacto social:** El acceso a los productos financieros es un derecho de los ciudadanos, se debe ser cauteloso con las conclusiones ya que el tratamiento de este tipo de instrumentos, impacta en el bienestar social.

Diego Lagos

Business Intelligence Analyst

[Linkedin](#)

+34 632 67 31 73

[diegoandres0905@gmail.com](mailto:diegoandres0905@gmail.com)