

Problema Empresarial

Una institución bancaria portuguesa enfrenta ineficiencias significativas en sus campañas de marketing directo telefónico para depósitos a plazo. Según datos históricos (2008-2010), se requieren múltiples contactos por cliente para determinar su interés, lo que genera costos operativos elevados y una baja productividad del equipo de telemarketing. La ausencia de una segmentación predictiva resulta en contactos masivos e indiscriminados, desperdiciando recursos en prospectos con baja probabilidad de conversión. Esto impacta directamente al departamento de marketing (ROI deficiente), al **centro de llamadas (call center)** (productividad reducida) y a la gestión comercial (metas de captación no alcanzadas), afectando así la estrategia de crecimiento de la cartera de productos de inversión del banco.

Justificación del Uso de Ciencia de Datos e IA:

Mediante el **Aprendizaje Automático (Machine Learning)**, podemos implementar un modelo predictivo que mejore la eficiencia, analizando los patrones en los datos categóricos de este **conjunto de datos (dataset)**. Paralelamente, utilizando modelos como la **regresión logística (logistic regression)**, concepto que estamos estudiando en el curso de **aprendizaje automático (machine learning)**, podríamos desarrollar un modelo para clasificar si un cliente contratará o no un depósito a plazo.

Formulación de Pregunta SMART:

¿Es posible crear un modelo de **Aprendizaje Automático (Machine Learning)**, utilizando el conjunto de datos bank-additional-full.csv, que prediga la suscripción a un depósito a plazo con una **exactitud (accuracy)** mínima de entre el 70 % y el 85 %? El objetivo es aplicar este modelo a un conjunto de 500 prospectos para alcanzar una tasa de conversión del 10 % al 30 %.

- **S (Específico - Specific):** Predecir qué clientes de un nuevo grupo de 500 prospectos se suscribirán a un depósito a plazo, identificando a aquellos con mayor probabilidad para optimizar el contacto.
- **M (Medible - Measurable):** Lograr una **exactitud (accuracy)** global del modelo de al menos el 85 % y una **precisión (precision)** de al menos el 70 % para la clase positiva (clientes que sí se suscriben). Se busca obtener una tasa de conversión del 10 % al 30 % en el grupo de prospectos seleccionados por el modelo.
- **A (Alcanzable - Achievable):** Con el conjunto de datos proporcionado, es factible alcanzar el rendimiento esperado del modelo.
- **R (Relevante - Relevant):** Este modelo puede tener un impacto significativo en las metas de la organización, aumentando la eficiencia del **centro de llamadas (call center)** y el retorno de la inversión.
- **T (Temporal - Time-bound):** El objetivo debe alcanzarse antes de que finalice el presente semestre académico.

Justificación técnica:

El contenido de las variables predictoras y el volumen de datos deberían permitir desarrollar los objetivos del proyecto de este curso en un tiempo razonable y con métricas medibles.

Referencias.

- Yamahata, H. (n.d.). *Bank Marketing* [Data set]. Kaggle. Recuperado el 16 de agosto de 2025 de <https://www.kaggle.com/datasets/henriqueyamahata/bank-marketing>. (Kaggle)
- Moro, S., Rita, P., & Cortez, P. (2014). *Bank Marketing* [Data set]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5K306>. (UCI Machine Learning Repository)
- Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. <https://doi.org/10.1016/j.dss.2014.03.001>. (scirp.org)