



www.aileaders-project.eu

EJERCICIO DE SIMULACIÓN -

**La importancia de la calidad de los datos en
las campañas de marketing dirigidas por IA**



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

EJERCICIO DE ESCENARIO

– La importancia de la calidad de los datos en las campañas de marketing dirigidas a la IA marketing

01



Resumen 3

02



Introducción 3

03



Presentación de herramientas 4

04



Actividades prácticas 5

05



Conclusiones 6

06



Referencias 7

07



Material complementario 7

• 01 Resumen



Tipo de REA

Ejercicio de simulación

Objetivo/Propósito

Concienciar sobre la importancia de la calidad de los datos para la implementación de algoritmos de toma de decisiones automatizada (ADMS), especialmente en el sector del marketing.

Resultados de aprendizaje esperados

El alumno será capaz de implementar medidas para abordar sesgos en las predicciones del comportamiento de los clientes.

Palabras clave

- Calidad de los datos
- Algoritmos sesgados
- Datos
- desequilibrados
- Campañas de marketing
- ADMS

Sugerido Metodológico Enfoque

Aprendizaje basado en problemas

• 02 Introducción



Contexto: predicción del marketing bancario

Éxito mediante el aprendizaje automático

- 01** El conjunto de datos utilizado en este proyecto es el conjunto de datos de marketing bancario del repositorio de aprendizaje automático de la UCI. Contiene detalles sobre los clientes contactados en una campaña de marketing y si se suscribieron a un depósito a plazo fijo.
- 02** Descripción del conjunto de datos: Conjunto de datos de marketing bancario – UCI (bank.csv incluido en los materiales OER).
- 03** Lea atentamente cada variable y comprenda su significado.

Enunciado del problema clave

- 01** ¿Cómo podemos predecir si un cliente suscribirá un depósito a plazo fijo basándonos en su perfil y en sus interacciones con campañas anteriores?
- 02** **Reto:** el conjunto de datos está muy desequilibrado, ya que son muchos menos los clientes que contratan un depósito, lo que puede dar lugar a un sesgo en el modelo.

• 03 Presentación de herramientas



Herramientas utilizadas en este ejercicio de simulación

01 Python

El lenguaje de programación más utilizado en ciencia de datos

02 Un editor de Python

Google Colab o Jupyter Notebook

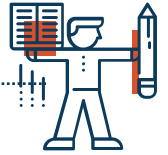
03 Bibliotecas y marcos

- Pandas: manipulación y limpieza de datos
- Scikit-learn: entrenamiento de modelos, evaluación y preprocesamiento de datos
- XGBoost: marco optimizado de refuerzo de gradiente
- imbalanced-learn (SMOTE): para abordar problemas de desequilibrio de clases
- Matplotlib/Seaborn: visualización y análisis de datos

04 Técnicas aplicadas

- Limpieza de datos e ingeniería de características
- Gestión del desequilibrio de clases (SMOTE, submuestreo, pesos de clases)
- Entrenamiento de modelos (RandomForest, XGBoost)
- Evaluación del rendimiento (recuerdo, precisión y puntuación F1)

• 04 Actividades prácticas



01 Preprocesamiento de datos

- Se eliminaron variables como la duración, la campaña y los días de pago, que solo se conocen tras contactar con los clientes, para evitar la filtración de datos.
- Codificación de variables categóricas y escalado de datos numéricos para mejorar el rendimiento del modelo.

02 Gestión del desequilibrio de clases

Aplicado:

- Ponderación de clases en RandomForest para penalizar los errores en la clase minoritaria.
- SMOTE (técnica de sobremuestreo sintético de minorías) para aumentar artificialmente las muestras en la clase minoritaria.
- Submuestreo para reducir el tamaño de la clase mayoritaria.

03 Entrenamiento del modelo

Se probaron varios modelos:

- RandomForest para la predicción de referencia.
- XGBoost para mejorar el rendimiento, optimizado con Early Stopping y Feature Selection para reducir el tiempo de entrenamiento.

04 Evaluación

Evaluación de los resultados utilizando

- Recall para la detección de clases minoritarias.
- F1-Score para una precisión equilibrada entre precisión y recall.



Ideas clave

01 El sesgo en los datos es importante:

El desequilibrio en el conjunto de datos llevó a los modelos iniciales a ignorar la clase minoritaria (clientes que se suscriben a depósitos).

02 Las técnicas de equilibrio son fundamentales:

El submuestreo, las ponderaciones de clase y SMOTE mejoraron significativamente la recuperación, aunque con compensaciones en la precisión general.

03 XGBoost con selección de características:

Al reducir el número de características y añadir la detención temprana, XGBoost mejoró el rendimiento sin comprometer la eficiencia.

Lecciones clave

01 Diseño ético de la IA:

El diseño ético de la IA requiere una preparación cuidadosa de los conjuntos de datos, métricas de evaluación justas y conciencia de los posibles sesgos en los resultados.

• 06 Referencias



- Conjunto de datos de marketing bancario - Repositorio de aprendizaje automático de la UCI: <https://archive.ics.uci.edu/dataset/222/bank+marketing>
- Documentación de Scikit-learn: <https://scikit-learn.org/>
- Documentación de XGBoost: <https://xgboost.readthedocs.io/en/stable/>
- Documentación de Imbalanced-learn: <https://imbalanced-learn.org/>



• 07 Material complementario



Archivo Jupyter Notebook (IPYNB)

Se incluye un cuaderno Python detallado con código comentado que guía a través de cada paso del proceso.

El cuaderno contiene

- Pasos de limpieza y preprocesamiento de datos.
- Ingeniería de características y lógica de selección de variables.
- Implementación de diferentes técnicas de equilibrio (SMOTE, pesos de clase y submuestreo).
- Entrenamiento del modelo con RandomForest y XGBoost.
- Métricas de evaluación y conclusiones a partir de los resultados.



Sigue nuestro viaje



www.aileaders-project.eu



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.