



www.aileaders-project.eu

EXERCÍCIO COM BASE EM CENÁRIO -

A importância da qualidade dos dados nas campanhas de marketing direcionadas 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.

EXERCÍCIO DE CENÁRIO

– A importância da qualidade dos dados nas campanhas de marketing direcionadas por IA

01



Resumo 3

02



Introdução 3

03



Apresentação das ferramentas 4

04



Atividades práticas 5

05



Conclusões 6

06



Referências 7

07



Material complementar 7

• 01 Resumo



Tipo de Recurso Educacional Aberto (OER)

Exercício com base em cenário

Objetivo/Finalidade

Sensibilizar para a importância da qualidade dos dados para a implementação de algoritmos de tomada de decisão automatizada (ADMS), particularmente no setor de marketing.

Resultados de aprendizagem esperados

O estudante será capaz de implementar medidas para lidar com enviesamentos nas previsões de comportamento dos clientes.

Palavras-chave

- Qualidade dos dados
- Algoritmos tendenciosos
- Dados
- desequilibrados
- Campanhas de marketing
- ADMS

Abordagem Metodológica sugerida:

Aprendizagem baseada em problemas

• 02 Introdução



Contexto: previsão do sucesso de marketing bancário usando machine learning

- 01** O conjunto de dados utilizado neste projeto é o conjunto de dados de Marketing Bancário do Repositório de Machine Learning da UCI. Ele contém detalhes sobre os clientes contactados numa campanha de marketing e se eles subscreveram um depósito a prazo.
- 02** Descrição do conjunto de dados: [Bank Marketing Dataset – UCI](#) (bank.csv incluído nos materiais de Recursos Educacionais Abertos (OER)).
- 03** Leia atentamente cada variável e compreenda o seu significado.

Declaração do problema principal

- 01** Como podemos prever se um cliente irá subscrever um depósito a prazo com base no seu perfil e nas interações com campanhas anteriores?
- 02** **Desafio:** O conjunto de dados é altamente desequilibrado, com muito menos clientes a subscreverem um depósito, o que leva a **um potencial viés no modelo.**

• 03 Apresentação das ferramentas



Ferramentas utilizadas neste exercício de cenário

01 Python

A linguagem de programação mais utilizada em Ciência de Dados

02 Um editor Python

Google Colab ou Jupyter Notebook

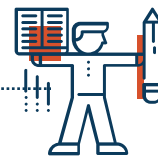
03 Bibliotecas e estruturas

- Pandas: manipulação e limpeza de dados
- Scikit-learn: treino de modelos, avaliação e pré-processamento de dados
- XGBoost: estrutura otimizada por reforço do gradiente
- Imbalanced-learn (SMOTE): para resolver problemas de desequilíbrio de classes
- Matplotlib/Seaborn: visualização e insights de dados

04 Técnicas aplicadas

- Limpeza de dados e engenharia de características
- Tratamento do desequilíbrio de classes (SMOTE, subamostragem, pesos de classe)
- Treino de modelos (RandomForest, XGBoost)
- Avaliação de desempenho (recall, precision e F1-score)

• 04 Atividades práticas



01 Pré-processamento de dados

- Variáveis removidas, como duração, campanha e pdays, que só são conhecidas após o contacto com os clientes, para evitar fuga de dados.
- Variáveis categóricas codificadas e dados numéricos escalonados para melhor desempenho do modelo.

02 Tratamento do desequilíbrio de classes

Aplicado:

- Ponderações de classe no RandomForest para penalizar erros na classe minoritária.
- SMOTE (Synthetic Minority Over-sampling Technique) para aumentar artificialmente as amostras na classe minoritária.
- Subamostragem para reduzir o tamanho da classe maioritária.

03 Treino do modelo

Vários modelos testados:

- RandomForest para previsão de linha de base.
- XGBoost para melhorar o desempenho, otimizado com Early Stopping e Feature Selection para reduzir o tempo de treino.

04 Avaliação

Resultados avaliados, usando

- Recall para deteção de classes minoritárias.
- F1-score para exatidão equilibrada entre precision e recall.



Principais insights

01 O viés nos dados é importante:

O desequilíbrio no conjunto de dados levou os modelos iniciais a ignorar a classe minoritária (clientes que subscrevem depósitos).

02 As técnicas de equilíbrio são fundamentais:

A subamostragem, os pesos de classe e o SMOTE melhoraram significativamente a recuperação, embora com compromissos na precisão geral.

03 XGBoost com seleção de características:

Ao reduzir o número de características e adicionar a interrupção precoce, o XGBoost melhorou o desempenho sem comprometer a eficiência.

Principais lições

01 Design ético de IA:

O design ético de IA requer uma preparação cuidadosa do conjunto de dados, métricas de avaliação justas e consciência do potencial viés nos resultados.

• 06 Referências



- Conjunto de dados de marketing bancário - Repositório de machine learning da UCI: <https://archive.ics.uci.edu/dataset/222/bank+marketing>
- Documentação Scikit-learn: <https://scikit-learn.org/>
- Documentação do XGBoost: <https://xgboost.readthedocs.io/en/stable/>
- Documentação do Imbalanced-learn: <https://imbalanced-learn.org/>



• 07 Material complementar



Ficheiro Jupyter Notebook (IPYNB)

Está incluído um Python Notebook detalhado com código comentado que orienta cada etapa do processo.

O notebook contém

- Etapas de limpeza e pré-processamento de dados.
- Engenharia de recursos e lógica de seleção de variáveis.
- Implementação de diferentes técnicas de equilíbrio (SMOTE, pesos de classe e subamostragem).
- Treino de modelos com RandomForest e XGBoost.
- Métricas de avaliação e insights a partir dos resultados.



leaders

Acompanhe a nossa jornada



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.