

Inteligência Artificial

Projeto 2

Portuguese  
Banking Institution  
Campaign

Bruna Brasil Marques up202007191  
Maria Sofia Gonçalves up202006927  
Pedro Miguel Correia up202006199  
3LEIC08 Grupo 88



# Portuguese Banking Institution Campaign

- Os dados são referentes a uma campanha de marketing de uma instituição bancária portuguesa que tinha como objetivo apurar que clientes é que subscreveriam a um determinado produto (um depósito bancário a prazo).
- A campanha de marketing foi baseada em telefonemas. Muitas vezes, foi estabelecido mais do que um contacto com o cliente para avaliar se iria subscrever(1) ou não(0) o produto.

# Bibliografia

- <https://www.kaggle.com/datasets/umanglodaya/portuguese-banking-institution-campaign>
- Biblioteca SKLearn.

# Ferramentas e algoritmos utilizados

- Até ao momento, foram usadas as bibliotecas de Python numpy, seaborn, pandas, matplotlib e sklearn.
- Os algoritmos utilizados foram neural networks, decision tree, knn (k-nearest neighbors).

# Trabalho desenvolvido

- Fizemos um pré-processamento dos dados:
  - Removemos colunas que tinham uma quantidade significativa de valores nulos. Caso esta quantidade não fosse grande, preenchemos os espaços vazios;
  - Passamos todas as colunas do tipo objeto para o tipo int64, para aplicarmos os algoritmos, com a ajuda do *LabelEncoder*;
  - Criámos uma nova coluna que representa se o dia em que foi realizada a chamada com o cliente foi na primeira ou segunda metade do mês.
- Criámos os ficheiros neuralNetworks.ipynb, decisionTree.ipynb e knn.ipynb, nos quais aplicámos os algoritmos de mesmo nome.

# Pré-processamento dos dados

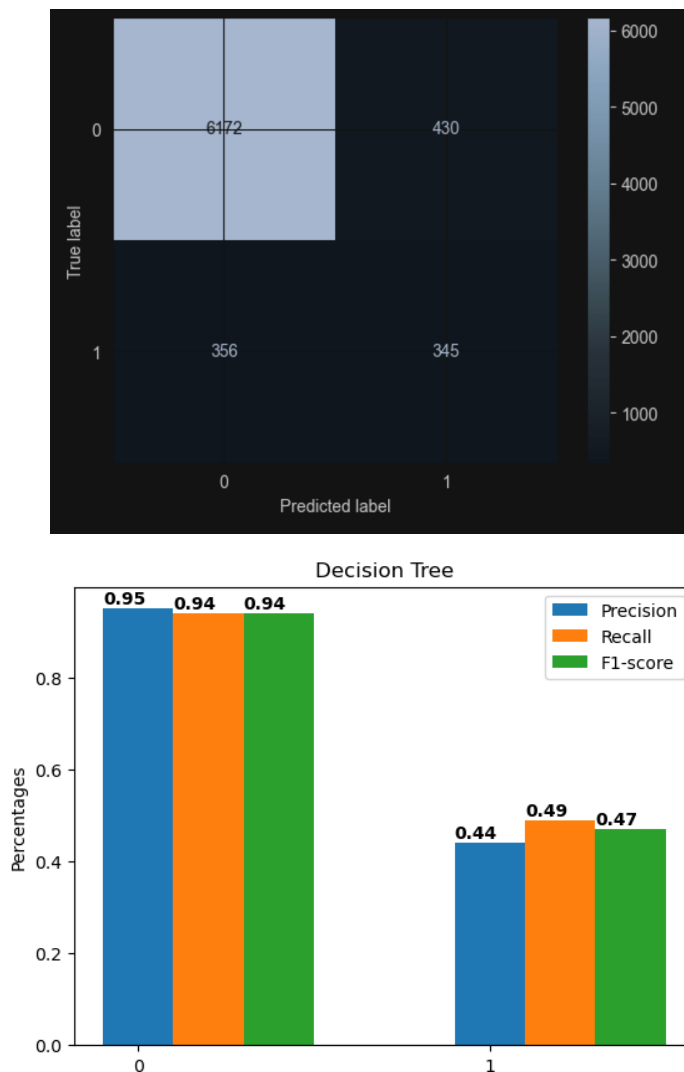
- Começámos por dar drop da coluna ID, uma vez que não tem qualquer interesse para o trabalho realizado (o ID é um número único para cada cliente, não afetaria se escolheu subscrever o produto do banco ou não).
- De seguida, fomos ver a informação da tabela: as colunas com valores nulos, a sua quantidade e o seu tipo. Depois, contabilizámos os valores nulos em cada coluna, dando drop às que tinham uma quantidade significativa deles. Nas colunas com um reduzido número de valores nulos, preenchemos esses valores com a média/moda dessa mesma coluna (dependendo do caso). Substituímos a coluna `day_of_month` por `mid_month`, ou seja, se o dia em que a pessoa foi contactada foi na primeira metade (0) ou na segunda metade (1) do mês. Por último, passámos todas as colunas do tipo Object para int32, com a ajuda do Label Encoder.

# Modelos desenvolvidos

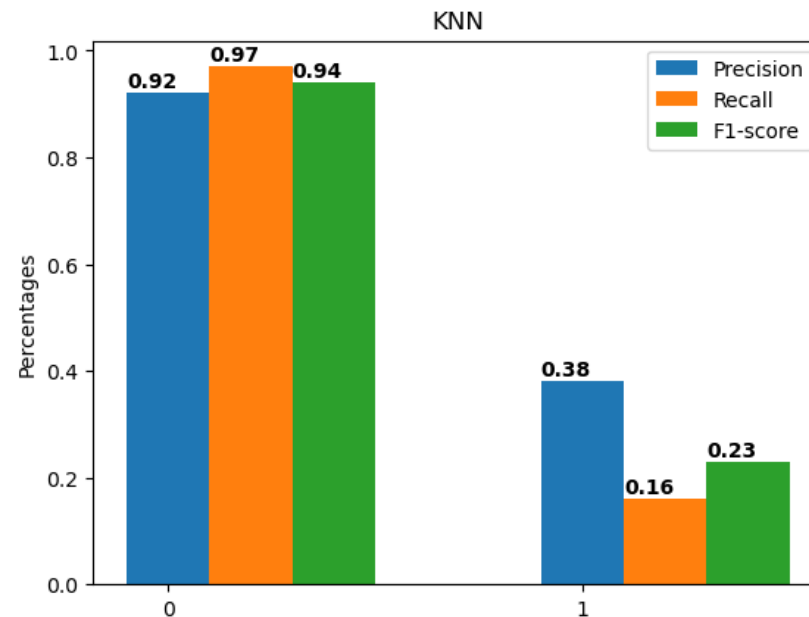
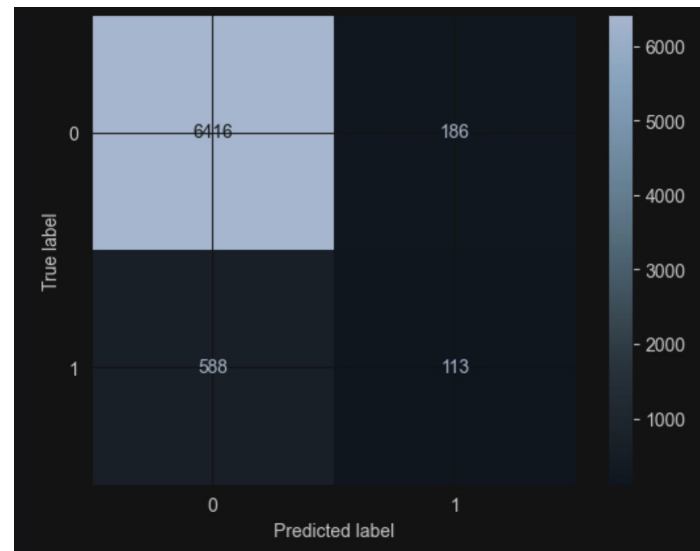
- **Decision Tree:** utiliza uma estrutura em forma de árvore para representar decisões e possíveis consequências. Para construir a árvore, o algoritmo procura dividir o conjunto de dados em subconjuntos. O resultado é uma decision tree que pode ser utilizada para classificar novos objetos.
- **K-Nearest Neighbours:** calcula a distância entre um novo dado e todos os dados já existentes, e os K dados mais próximos são utilizados para determinar a classe do novo dado, a partir da classe mais comum entre os K vizinhos.
- **Neural Networks:** tenta imitar o funcionamento do cérebro humano, consistindo num conjunto de neurónios conectados que processam informações para produzir um output. O processo de treino de uma neural network envolve a apresentação de um conjunto de exemplos de entrada (input) juntamente com as saídas (output) esperadas. A rede é ajustada para minimizar a diferença entre as saídas esperadas e as saídas produzidas pela rede. Esse processo é repetido várias vezes até que a rede produza saídas precisas o suficiente.

# Avaliação individual

- Decision Tree



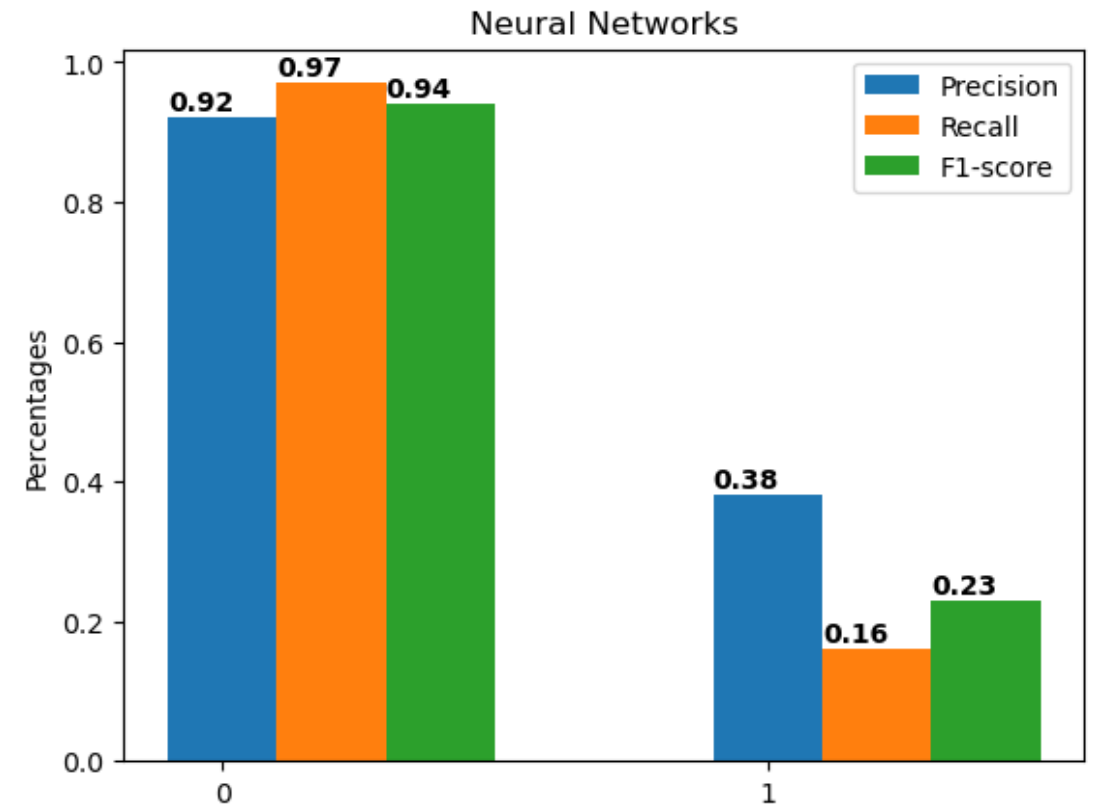
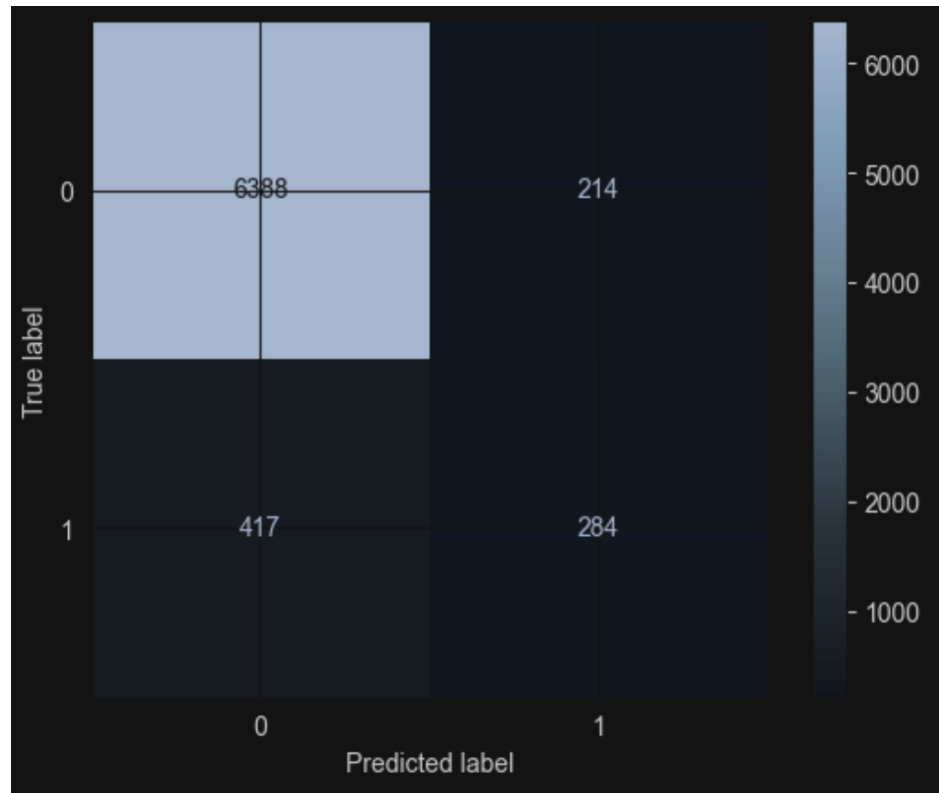
- K-Nearest Neighbours





# Avaliação individual

- Neural Networks



# Avaliação e Comparação

Como é possível observar pelas barras resultantes dos valores de accuracy de cada algoritmo, têm todos mais ou menos o mesmo valor, tirando o algoritmo de Neural Networks, que têm um valor ligeiramente superior.

Algoritmo	Decision Tree	KNN	Neural Networks
Tempo (s)	0.16	0.05	65.83

