

## **Estudos de Caso**

- 1) **O estudo de caso 1** envolve a aplicação prática de um problema de Ciência de Dados relacionado a Machine Learning. Você vai abordar o problema de Detecção de Transações Fraudulentas, utilizando o conjunto de dados disponível em Fraud Transaction Detection (<https://www.kaggle.com/code/llabhishekl/fraud-transaction-detection/input>) . O objetivo é desenvolver um modelo que seja capaz de prever se uma transação é fraude ou não. Você pode se inspirar no notebook presente em (<https://www.kaggle.com/code/llabhishekl/fraud-transaction-detection/notebook>). Para este problema de dados, você deve realizar as seguintes etapas:
- a) Elabore a análise exploratória da base de dados - análise gráfica e por tabelas. (25%)
  - b) Realize a previsão por modelos criados utilizando os algoritmos trabalhados em sala de aula, KNN, Árvore de Decisão e Regressão Logística. Compare os resultados.(40%)
  - c) Encontre as features mais importantes para a detecção de fraudes, interprete os resultados. (10%)
  - d) Apresente a matriz de confusão, acurácia, precisão e recall (15%)
  - e) Realize ajustes a fim de melhorar o desempenho do modelo e apresente os resultados. (10%)

Lembre-se de fornecer explicações detalhadas e apresentar os cálculos, gráficos e conclusões de forma clara e organizada.

Você deve anexar o notebook utilizado.

2) **O estudo de caso 2** envolve a aplicação prática de um problema de Ciência de Dados relacionado a Machine Learning. Você vai abordar o problema de Previsão de Cancelamento de Reservas em Hotéis, utilizando o conjunto de dados disponível em Hotel Booking (<https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand/data>) . O objetivo é desenvolver um modelo que seja capaz de prever se os indivíduos irão ou não cancelar suas reservas em uma rede de hotéis. Para este problema de dados, você deve realizar as seguintes etapas:

- a) Elabore a análise exploratória da base de dados - análise gráfica e por tabelas. (25%)
- b) Realize a previsão por um modelo de Regressão Logística (40%)
- c) Encontre as features mais importantes para o cancelamento das reservas, interprete os resultados. (10%)
- d) Apresente a matriz de confusão, acurácia, precisão e recall (15%)
- e) Realize ajustes a fim de melhorar o desempenho do modelo e apresente os resultados. (10%)

Lembre-se de fornecer explicações detalhadas e apresentar os cálculos, gráficos e conclusões de forma clara e organizada.

Você deve anexar o notebook utilizado.