



**Universidade Federal de Alagoas**  
**Instituto de Computação**  
**Ciência de Dados**



Professor: Bruno Pimentel

Aluno: \_\_\_\_\_

**Lista de Exercícios 1**

**1. Aquisição de Dados e Leitura**

Escreva um script em Python que leia um arquivo CSV contendo dados de vendas de uma empresa. O arquivo deve conter as seguintes colunas: **Data**, **Produto**, **Quantidade**, **Preço**. Utilize a biblioteca **pandas** para ler o arquivo e exibir as primeiras 5 linhas do DataFrame resultante.

**2. Pré-processamento de Dados**

Limpe os dados lidos no exercício anterior removendo linhas com valores nulos e convertendo a coluna **Data** para o tipo **datetime**. Em seguida, normalize a coluna **Quantidade** para que os valores estejam entre 0 e 1.

**3. Análise Estatística**

Calcule a média, mediana, desvio padrão e moda para a coluna **Preço** do DataFrame resultante do pré-processamento.

**4. Visualização de Dados**

Crie um gráfico de barras utilizando a biblioteca **matplotlib** ou **seaborn** que mostre a quantidade total vendida de cada produto. O eixo x deve representar os produtos e o eixo y a quantidade total vendida.

**5. K-Vizinhos Mais Próximos - Classificação**

Usando a mesma base de dados, crie uma coluna binária chamada **Alta\_Venda**, onde o valor é 1 se a quantidade vendida for maior que a média e 0 caso contrário. Em seguida, crie um modelo de classificação utilizando o algoritmo K-Vizinhos Mais Próximos (KNN) para prever se uma venda será alta ou não. Avalie o modelo utilizando a matriz de confusão.

## **6. Agrupamento - K-means**

Utilize o algoritmo K-means para agrupar os produtos com base em suas quantidades vendidas e preços. Determine o número ideal de clusters utilizando o método do cotovelo (elbow method).

## **7. Análise de Clusters**

Descreva as características de cada cluster encontrado no exercício anterior. Identifique quais produtos estão em cada cluster e discuta possíveis razões para esses agrupamentos.

## **8. Visualização de Clusters**

Crie uma visualização que mostre os clusters formados pelo algoritmo K-means. Utilize um gráfico de dispersão, onde cada ponto representa um produto, e cores diferentes representam os diferentes clusters.

## **9. Validação Cruzada**

Realize uma validação cruzada de 5 vezes para o modelo KNN criado no exercício 5. Compare os resultados obtidos e discuta a estabilidade do modelo.

## **10. Tomada de Decisão baseada em Clustering**

- Com base nos clusters identificados, escreva um script que sugira estratégias de marketing. Por exemplo, para cada cluster, decida se é melhor aumentar ou diminuir o preço dos produtos, lançar promoções ou focar em marketing específico.
- Aplique essas mudanças aos dados históricos e compare os resultados antes e depois da aplicação das estratégias usando métodos estatísticos como teste t, teste de Wilcoxon, ou análise de variância (ANOVA) para avaliar se as mudanças implementadas tiveram um impacto significativo nas vendas.
- Gere um relatório resumindo os resultados da análise estatística e discuta se as estratégias de marketing sugeridas foram eficazes.