

Prática 7 – Agrupamento dos Dados de Teste do Trabalho Prático com K-Means

Objetivo: Aplicar técnicas de **Aprendizado Não Supervisionado**, utilizando o algoritmo **K-Means**, para identificar agrupamentos nos **dados de teste** utilizados no trabalho prático.

Data de Entrega e Apresentação: 18/06/2025

Tarefa:

- Carregue o conjunto de dados de teste utilizado no seu trabalho prático. Verifique se há necessidade de tratamento de valores ausentes ou normalização dos atributos numéricos.
- Selecione os atributos relevantes para a formação dos grupos. Explique sua escolha dos atributos com base na intuição ou conhecimento do domínio. Evite atributos altamente correlacionados ou identificadores.
- Normalize os dados utilizando *StandardScaler* ou método equivalente.
- Aplique o algoritmo K-Means, iniciando com $k = 3$. Visualize os agrupamentos usando um gráfico de dispersão (scatter plot), se possível. Obtenha os **centróides** dos grupos e crie um **gráfico de barras agrupadas** comparando os valores médios por atributo.
- Repita o experimento com $k=2$ e $k=4$. Compare os agrupamentos obtidos com os de $k=3$. Escolha o valor de k que, na sua avaliação, gera o agrupamento mais coerente.
- Analise os grupos formados:
 - Quais características são comuns entre os elementos de cada grupo?
 - Houve separação clara?
 - Algum grupo parece conter *outliers*, ou seja, instâncias que se desviam significativamente do padrão geral dos outros dados?

Análise:

- Como a escolha do valor de k afetou os resultados?
- Os agrupamentos fazem sentido em termos práticos ou de negócio?
- Como o K-Means lidou com os dados do seu trabalho prático?
- Os agrupamentos poderiam ser utilizados para melhorar de alguma forma o resultado da classificação?

Entrega: A dupla deverá entregar e apresentar:

- Um notebook Jupyter (.ipynb) contendo:
 - O código-fonte do agrupamento.
 - As visualizações dos grupos.
 - O gráfico de barras dos centróides.
 - Uma caixa de texto breve com análise solicitada acima.