# Aprendizado de Máquina

2024.1

Atividade da segunda unidade

Objetivo

O objetivo deste trabalho é aplicar técnicas de clustering (agrupamento) para explorar uma base de dados, identificar padrões e tirar conclusões interessantes sobre os dados. Devem ser utilizados diferentes algoritmos de clustering e avaliar o desempenho de cada método.

## Base de Dados

Para este trabalho, os alunos utilizarão a base de dados “Mall Customers”, disponível no repositório do Kaggle: <https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python>

A base de dados contém informações sobre clientes de um shopping, incluindo variáveis como:

* CustomerID: Identificação única do cliente
* Gender: Gênero do cliente
* Age: Idade do cliente
* Annual Income: Renda anual em milhares de dólares
* Spending Score: Pontuação de gastos atribuída pelo shopping, baseada no comportamento do cliente e na natureza dos gastos

## Instruções do Trabalho

### Preparação dos Dados:

* Carregar a base de dados e realizar uma análise exploratória inicial.
* Tratar valores ausentes, se houver, e realizar a normalização dos dados, se necessário.
* Visualizar a distribuição dos dados utilizando gráficos apropriados.

### Algoritmos de Agrupamento:

* Implementar pelo menos dois algoritmos de agrupamento diferentes, por exemplo:
  + K-Means
  + Agglomerative Clustering
* Explicar brevemente o funcionamento de cada algoritmo e os parâmetros utilizados.

### Avaliação dos Resultados

* Utilizar métricas de avaliação para analisar a qualidade dos clusters formados, como:
  + Silhouette Score
  + Davies-Bouldin Index
* Visualizar os clusters utilizando gráficos, como scatter plots em 2D ou 3D, se aplicável.

### Análise e Conclusões:

* Interpretar os resultados obtidos a partir dos clusters formados.
* Discutir as características dos clusters, identificando perfis de clientes e possíveis insights comerciais.
* Comparar os resultados dos diferentes algoritmos e discutir qual deles apresentou a melhor performance e por quê.

## Entrega

### Relatório

* Elaborar um relatório final contendo:
  + Introdução ao problema e à base de dados
  + Metodologia utilizada para o agrupamento
  + Resultados obtidos com cada algoritmo
  + Análise e discussões dos resultados
  + Conclusões e possíveis aplicações dos insights encontrados

### Código

* O trabalho deverá ser entregue em formato Jupyter Notebook, contendo todo o código e visualizações.
* O relatório final deverá ser entregue em formato PDF.

### Prazo e Meio

* Prazo para entrega: 26/07
* Meio para entrega: atividade no SIGAA
* **Trabalho pode ser feito em grupos de até duas pessoas.**

## **Critérios de Avaliação**

* Completude e qualidade da análise exploratória dos dados.
* Implementação/utilização correta e funcionamento dos algoritmos de agrupamento.
* Qualidade das visualizações e das análises dos resultados.
* Clareza e profundidade das conclusões e discussões.
* Organização e apresentação do relatório final.

**Nota do projeto: 5,0 pontos na segunda unidade**

## Observação

Não é necessário implementar os algoritmos de agrupamento. Pacotes de implementação, como o Scikit-Learn (ou outros) podem ser utilizados livremente.