

## OBS: Projeto Individual

**Dia da apresentacao: 26/11/2025**

**Envio dos slides até: 25/11/2025**

# 1 Problema

O objetivo deste trabalho é selecionar um dataset de classificação disponível no *OpenML* (<https://www.openml.org/>) e realizar a avaliação de diferentes algoritmos de aprendizado de máquina supervisionado.

O banco de dados escolhido deve atender aos seguintes critérios:

- Conter mais de 10 atributos (variáveis preditoras);
- Ter mais de 1000 instâncias (amostras).

Cada um deverá informar antecipadamente, por meio de formulário (<https://forms.gle/X5hF8WUqpFBdXSXx9>), qual conjunto de dados será utilizado. Os classificadores a serem avaliados incluem, no mínimo, os seguintes algoritmos:

- K-Nearest Neighbors (kNN);
- Perceptron;
- MLP(Multi Layer Perceptron);
- Naive Bayes;

A comparação de desempenho será feita por meio das métricas:

- *Accuracy*;
- *F1-score*.

## 2 Estrutura Mínima do Trabalho

O projeto deve conter, no mínimo, as seguintes seções:

1. **Introdução**
2. **Algoritmos de Aprendizagem de Máquina**
3. **Experimentos**
  - 3.1. **Banco de Dados**
  - 3.2. **Métricas de Avaliação**
  - 3.3. **Resultados**
4. **Conclusões**
5. **Referências**

## 3 Classificadores

- (a) **k-Nearest Neighbors (kNN)**: treine o classificador utilizando diferentes medidas de distância:
  - Distância Euclidiana;
  - Distância Manhattan;
- (b) **Perceptron**: implemente o modelo de Perceptron. Treine o modelo por múltiplas épocas.
- (c) **MLP (Multi-Layer Perceptron)**: implemente uma rede neural multicamada com, pelo menos, uma camada oculta.
- (d) **Naive Bayes**: implemente as versões univariada e multivariada do classificador bayesiano.

Calcule a média e o desvio padrão das métricas de avaliação (Acurácia e F1-Score), considerando todos os *folds* da validação cruzada. Em seguida, apresente os resultados em uma tabela comparativa, conforme o modelo abaixo.

### Exemplo de Tabela com Médias, Desvio Padrão e Tempos

Classificador	Acurácia	Precisão	F1-Score	Tempo Treino (s)	Tempo Teste (s)
K-vizinhos (Dist. Euclidiana)	$0.85 \pm 0.03$	$0.84 \pm 0.04$	$0.84 \pm 0.04$	$0.12 \pm 0.01$	$0.03 \pm 0.00$
K-vizinhos (Dist. Manhattan)	$0.88 \pm 0.02$	$0.87 \pm 0.03$	$0.87 \pm 0.03$	$0.11 \pm 0.01$	$0.02 \pm 0.00$
... Bayesiano (Univariado)	$0.87 \pm 0.03$	$0.86 \pm 0.03$	$0.86 \pm 0.03$	$0.07 \pm 0.01$	$0.01 \pm 0.00$

Tabela 1: Análise comparativa do desempenho dos classificadores, com médias, desvios padrão ( $\pm$ ) e tempos de treino e teste.

### Análise Comparativa dos Classificadores

Com base na Tabela 1, elabore uma análise crítica considerando:

- Comparação das métricas entre os classificadores (Acurácia e F1-Score);
- Diferenças nos tempos de treino e teste;
- Relação entre desempenho e eficiência computacional;
- Pontos fortes e limitações observadas em cada algoritmo.

Apresente, ao final, qual classificador apresentou melhor equilíbrio entre desempenho e tempo de execução para o problema analisado.

## Observações Importantes

- A entrega das implementações é **obrigatória**. Trabalhos sem código entregue receberão **nota zero**.
- A pontualidade será um critério avaliativo.
- Todos os trabalhos serão submetidos a ferramentas de detecção de plágio (incluindo IA generativa como ChatGPT, Copilot e DeepSeek). Casos confirmados resultarão em **nota zero**.
- Não é permitido o uso de bibliotecas que implementem diretamente os algoritmos (por exemplo, `scikit-learn`) nem o uso de `pandas`. Todo o código deve ser implementado manualmente.
- Os slides da apresentação deverão ser enviados no AVA dentro do prazo estabelecido. Devem conter:
  - Introdução e justificativa do dataset escolhido;
  - Explicação do código e funcionamento dos algoritmos;
  - Resultados e comparações;
  - Conclusões gerais.
- Não é necessário entregar um relatório escrito além dos slides.
- Caso algum aluno não consiga apresentar, deverá agendar uma avaliação individual. Caso contrário, será atribuída **nota zero**.