

Instruções: Entregar via e-mail em Google Colab ou similares. Não se esqueça de abrir uma caixa de texto no começo do arquivo e inserir um link do seu trabalho

---

## Trabalho 7: Data de Entrega 26/06/2025 até às 23h59min

---

1. O conjunto de dados **Heart Disease Dataset** da UCI Machine Learning Repository é bastante utilizado para estudos de predição de doenças cardíacas. Ele reúne informações médicas de pacientes para identificar se há ou não a presença da doença cardíaca. O arquivo está disponível na página do curso, na aba **Base de Dados**.

### Características do Dataset

- **Número de exemplos:** 303 pacientes.
- **Número de variáveis:** 14 variáveis explicativas e 1 variável alvo.
- **Objetivo:** Classificar se o paciente tem (1) ou não tem (0) doença cardíaca.

### Variáveis

- (a) **age** – Idade (anos).
- (b) **sex** – Sexo (1 = masculino, 0 = feminino).
- (c) **cp** – Tipo de dor no peito:
  - 1: angina típica
  - 2: angina atípica
  - 3: dor não anginosa
  - 4: assintomático
- (d) **trestbps** – Pressão arterial em repouso (mm Hg).
- (e) **chol** – Colesterol sérico (mg/dl).
- (f) **fbs** – Açúcar no sangue em jejum  $> 120$  mg/dl (1 = sim, 0 = não).
- (g) **restecg** – Resultado do eletrocardiograma em repouso:
  - 0: normal
  - 1: anormalidade ST-T
  - 2: hipertrofia ventricular esquerda
- (h) **thalach** – Frequência cardíaca máxima alcançada.
- (i) **exang** – Angina induzida por exercício (1 = sim, 0 = não).
- (j) **oldpeak** – Depressão do segmento ST induzida pelo exercício.
- (k) **slope** – Inclinação do segmento ST durante o esforço:
  - 1: ascendente

- 2: plano
  - 3: descendente
- (l) **ca** – Número de vasos principais coloridos por fluoroscopia (0 a 3).
- (m) **thal** – Resultado do exame de talassemia:
- 3: normal
  - 6: defeito fixo
  - 7: defeito reversível
- (n) **target** – Presença (1) ou ausência (0) de doença cardíaca.

Utilize o conjunto de dados para treinar modelos que façam a classificação da presença ou ausência da doença cardíaca. Para isso:

- Treine os seguintes algoritmos de classificação: **Regressão Logística**, **k-NN**, **SVM** e **Árvore de Decisão**.
- Utilize validação cruzada 5-fold para escolher os melhores parâmetros otimizados de cada modelo.
- Avalie o desempenho usando métricas da matriz de confusão (como acurácia, sensibilidade, especificidade) e a **área sob a curva ROC (AUC)**.
- Compare os modelos e escolha o que apresentar melhor desempenho para essa tarefa, justificando sua escolha.