

Exercício Avaliativo - Aplicação de Aprendizado de Máquina

Disciplina: Ciência de dados

Valor: 15 pontos

Data entrega: 05/06/24

Observação: O trabalho deverá ser realizado **individualmente**. Todos os códigos serão analisados detalhadamente, o que pode levar algum tempo para a correção. Por isso, **as notas podem demorar um pouco para serem divulgadas**.

1. Escolha um dataset

Escolha um dos seguintes datasets disponíveis publicamente:

- Iris Dataset (Flores de Íris): <https://archive.ics.uci.edu/dataset/53/iris>
- Titanic: Machine Learning from Disaster (Kaggle): <https://www.kaggle.com/c/titanic/data>
- Wine Dataset (Classificação de vinhos): <https://archive.ics.uci.edu/dataset/109/wine>
- Heart Disease UCI: <https://archive.ics.uci.edu/dataset/45/heart+disease>

2. Escolha um algoritmo de aprendizado supervisionado

Utilize um dos seguintes algoritmos:

- Regressão Logística
- k-Nearest Neighbors (kNN)
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)

(É permitido usar bibliotecas como scikit-learn, pandas e matplotlib.)

3. Divida o dataset

Separe os dados em treino (70%) e teste (30%).

4. Treine e teste o modelo

Treine o modelo com os dados de treino e avalie com os dados de teste.

Responde as perguntas a seguir, elas são obrigatórias para entrega:

1. Qual dataset você escolheu e por quê?

2. Qual algoritmo você utilizou? Justifique a escolha.
3. Como foi feita a preparação dos dados?
 - Tratamento de valores ausentes?
 - Normalização?
 - Codificação de variáveis categóricas?
4. Qual foi a acurácia (ou outra métrica adequada) do seu modelo no conjunto de teste?
5. Você fez algum ajuste de hiperparâmetros? Como ele afetou os resultados?