Disciplina: SCC0530 - Inteligência Artificial
Trabalho: Projeto de Aprendizado de Máquina

---

1. Objetivo

Este trabalho tem como objetivo proporcionar ao aluno a elaboração de um projeto de aprendizado de máquina utilizando alguns dos mais tradicionais algoritmos de classificação.

2. Sobre o Projeto

O projeto deve contemplar as seguintes etapas:

- Selecionar o dataset para a tarefa de classificação (**complexidade do dataset será objeto de avaliação**). Repositórios para encontrar datasets:
  - https://archive.ics.uci.edu/ml/datasets.php
  - https://www.kaggle.com/datasets
  - https://research.google/tools/datasets/
- Realizar os pré-processamentos necessários (tratamento de dados faltantes, seleção de atributos, padronização, etc.) no dataset selecionado;
- O experimento deverá ser feito utilizando validação cruzada (*10-fold cross validation*);
- Aplicar os seguintes algoritmos de classificação com diferentes configurações (hiperparâmetros): K-Nearest Neighbors (KNN), Naive Bayes, Árvores de Decisão e Multilayer Perceptron (MLP) (Não é necessário desenvolver os algoritmos, somente aplicá-los conforme exemplos disponibilizados);
- Utilizar a medida de avaliação mais adequada ao dataset selecionado.

Observação: A implementação do projeto deve ser feito em linguagem Python, e as seguintes bibliotecas serão de grande utilidade no desenvolvimento do mesmo:

- **Pandas**: manipulação e análise de dados.
- **Seaborn**: gráficos.
- **Scikit-learn**: algoritmos de classificação, medidas de avaliação, algoritmos de pré-processamento, validação cruzada, *grid search*, etc.

3. Relatório

O relatório será no formato de notebook (**jupyter notebook ou google colab**) e deverá conter as seguintes informações:

- Descrição do dataset selecionado;
    - Recomendável utilização de visualização de dados para melhor compreensão do dataset;
- Descrição resumida das técnicas utilizadas e dos algoritmos utilizados no projeto;
- Código comentado;
- Explicações das decisões tomadas no projeto;
- Resultados;
- Conclusão;
- É necessário desenvolver o relatório com análises e explicações.
    - **Relatório apenas com o código e gráficos soltos não serão bem avaliados.**

4. Apresentação

- Slides com as seguintes informações: descrição do dataset, pré-processamentos utilizados, resultados obtidos e conclusão;
- O tempo máximo da apresentação é de **12 minutos + 3 minutos de arguição**.

5. Informações importantes

- Este trabalho deverá ser desenvolvido por **grupos de 4 alunos.**
- Ao final do projeto um representante do grupo irá entregar no escaninho o relatório nos seguintes formatos: .ipynb (formato do notebook executável) e os slides da apresentação.
- **Importante: A complexidade do *dataset* selecionado também será objeto de avaliação.**
- Os conhecimentos necessários para o desenvolvimento do projeto serão adquiridos nas aulas expositivas.
- Serão disponibilizados alguns materiais auxiliares para o desenvolvimento do projeto.
- Monitoria (agendar horário de dúvidas com o monitor por meio do email: nicolas.rsantos1@gmail.com)
- O preenchimento dos grupos será feito no calendário de apresentações, pelo seguinte link:

  https://docs.google.com/spreadsheets/d/1xvAF9ILAfHxW0iOd0O4wcl04PyZkSZPW3eS4O3mi1PQ/edit?usp=sharing

6. Datas

- **Depósito do projeto** no escaninho dos arquivos citados na Seção 5 **até às 23:59 do dia 20 de Junho de 2023.** Para cada dia de atraso será descontado 1 ponto na nota final.