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

---

1. Objetivo

Este projeto 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

Este projeto deverá contemplar as seguintes etapas:

- Selecionar dez *datasets*.
- Para cada *dataset* aplicar os pré-processamentos necessários (tratamento de dados faltantes, seleção de atributos, padronização, etc.).
- Selecionar a medida(s) de avaliação mais adequada aos *datasets* selecionados.
- Utilização de validação cruzada (*10-fold cross validation*) nos experimentos.
- 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).
- Para a análise dos resultados selecionar os dois melhores algoritmos de cada dataset e, dentre os dois, verificar o melhor por meio da diferença absoluta conforme os slides de aula.
- A avaliação do projeto será por meio de relatório e apresentação.

3. Sobre o Relatório

O relatório será no formato de notebook e deverá conter as seguintes informações:

- Descrição dos datasets selecionados e a tarefa de classificação relacionada ao dataset. Recomendável utilização de visualização de dados para melhor compreensão dos *datasets* utilizados.
- Descrição resumida das técnicas utilizadas e dos algoritmos utilizados no projeto.
- Código comentado.
- Explicações das decisões tomadas no projeto.
- Conclusão.

4. Sobre a apresentação

O objetivo da apresentação é mostrar diversas possibilidades de aplicações do uso de algoritmos de classificação para a turma. A apresentação deve conter:

- Descrição breve dos *datasets* selecionados.
- Decisões sobre os pré-processamentos utilizados.
- Resultados.

5. Informações importantes

- Este trabalho deverá ser desenvolvido por **grupos de 3 alunos.**
- A definição do grupo, dia de apresentação e ordem de apresentação devem ser registradas na seguinte planilha:
    - https://docs.google.com/spreadsheets/d/1TeYCteT-wzUawUVtME_1YhMwiPsdvUhVvCVjnKvuYXk/edit?usp=sharing
- Repositórios para encontrar datasets.
    - https://archive.ics.uci.edu/ml/datasets.php
    - https://www.kaggle.com/datasets
    - https://research.google/tools/datasets/
- A implementação do projeto será 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.
- Desenvolver o projeto em formato de notebook (jupyter notebook ou google colab)
- Ao final do projeto um representante do grupo irá entregar no escaninho o relatório nos seguintes formatos: .ipynb (formato do notebook executável, para ser executado em caso de dúvidas na avaliação), arquivo .html (que será o objeto de avaliação) e o slide da apresentação.
- O tempo de **apresentação** será de **15 minutos**.

6. Sobre dúvidas do projeto

- Os conhecimentos necessários para o desenvolvimento do projeto serão adquiridos nas aulas expositivas.
- Os horários de monitoria serão disponibilizados no Avisos do Tidia.
- Serão disponibilizados alguns materiais auxiliares para o desenvolvimento do projeto.

7. Datas

- **Depósito do projeto** no escaninho dos arquivos citados na Seção 5 **até às 23:59 do dia 21/06.** Para cada dia de atraso será descontado 1 ponto na nota final.
- **Apresentações nos dias 21/06, 24/06, 28/06 e 01/07. Link para preenchimento das datas:**
    - https://docs.google.com/spreadsheets/d/1TeYCteT-wzUawUVtME_1YhMwiPsdvUhVvCVjnKvuYXk/edit?usp=sharing