



Projeto Final

Descrição

O projeto final visa reforçar e aplicar os diversos conceitos e técnicas estudadas ao longo da disciplina, além de capacitar o aluno a apresentar, de forma detalhada e demonstrando conhecimento, as informações e os resultados encontrados com a utilizações das técnicas utilizadas.

Neste projeto, a ser realizado individualmente, você deverá:

1. Navegar no repositório de bases de dados da University of California Irvine (UCI), disponível neste endereço:
 - <https://archive.ics.uci.edu/ml/datasets.php>
2. Escolher um (1) dataset cuja tarefa principal seja aplicada à regressão **ou** classificação. O dataset escolhido deverá:
 - ser composto de pelo menos 5 atributos/features,
 - ter pelo menos 200 instâncias,
 - e, no caso da escolha ser feita para tarefas de classificação, o dataset deverá lidar com pelo menos 3 classes distintas.
3. Desenvolver, em Python, uma aplicação utilizando regressão múltipla ou regressão logística (utilizando a biblioteca de sua preferência) para lidar com o problema de regressão/classificação escolhido.
4. Escrever um artigo/relatório com os seguintes tópicos:
 - Introdução
 - Descrever o problema (não o conjunto de dados), fornecendo uma motivação para investigar o problema escolhido.
 - Materiais e Métodos
 - Descrever o conjunto de dados com informações numéricas e gráficas das informações disponíveis no dataset. Ex.: Apresente

(sempre de forma comentada!) figuras que ilustram o dataset, as features, e a relação entre as features ou classes contidas na base.

- Lembre-se que a visualização dos dados é uma boa estratégia para te ajudar a conhecer o problema e encontrar estratégias que podem melhor ajudar a resolver o problema.
- Descrever a solução adotada e o algoritmo desenvolvido, com detalhes e comentários para as escolhas feitas durante o processo de desenvolvimento.
 - Por exemplo: como funciona, de forma geral, o algoritmo utilizado para solução do problema? houve normalização dos dados? todas as features disponíveis na base de dados foram utilizadas na solução? você utilizou as configurações padrões do método disponível na biblioteca utilizada, ou alterou alguns parâmetros?
- Apresentar as escolhas realizadas para divisão de dados em conjuntos de treino e teste, quantidade de épocas (se for o caso) e quantidade de execuções do algoritmo para gerar o resultado que será comentado na seção seguinte.
 - Tenha certeza de incluir todas as informações necessárias para que um leitor, com sua explicação e informações, seja capaz de replicar todo o seu experimento obtendo o mesmo resultado que você.



- Resultados
 - Fornecer uma análise para os resultados obtidos. Apresente informações numéricas e/ou ilustrativas que fomentem ou justifiquem os seus comentários.
 - Nesta seção, você pode, por exemplo, apresentar taxas de acerto/erro para diferentes divisões de treino e teste, diferentes combinações de features utilizadas pelo algoritmo, utilização ou não de normalização dos dados, e o impacto dessas escolhas no resultado final.
- Conclusões
 - Fornecer as conclusões ou considerações finais sobre o trabalho desenvolvido.
- Referências
 - Apresentar a bibliografia utilizada para apoiar seu trabalho (livros, artigos, sites).
 - Procure referenciar as citações nos locais adequados, ao invés de simplesmente fornecer o link/referência utilizado.

O que deverá ser submetido:

- Um Python notebook (.ipynb) com o código desenvolvido, com cada etapa organizada e comentada de forma clara, e com o resultado da execução de cada célula visível (quando for o caso).
- Um arquivo .pdf com o artigo elaborado conforme instruções acima.

Data limite de entrega: até 23h59 do dia 05/06/2023, via google classroom