

Projeto AM 2022-2

Francisco de A. T. de Carvalho¹

1 Centro de Informatica-CIn/UFPE

Av. Prof. Luiz Freire, s/n -Cidade Universitaria, CEP 50740-540, Recife-PE, Brasil,
fatc@cin.ufpe.br

Questão 1

- Considere os dados "Avila Data Set" (Data set e artigo relevante em anexo).
- Use a ferramenta scikit learn (<https://scikit-learn.org/stable/>) ou a ferramenta WEKA (<https://www.cs.waikato.ac.nz/ml/weka/>) e use os 4 (quatro) classificadores abaixo:
 - i) Árvores de decisão
 - ii) Bayesiano ingenuo
 - iii) Regressão logística
 - iv) k-vizinhos. Use a distância Euclidiana para definir a vizinhança. Use conjunto de validação para fixar o número de vizinhos k .
- Treine o classificador no data set avila-tr.txt. Teste o classificador no data set avila-ts.txt. O conjunto de teste não pode ser usado no treinamento. Tentar melhorar cada um dos classificadores escolhido, procurando os melhores hiperparâmetros para alcançar a melhor precision (precisão), cobertura (recall) e f1-score no conjunto de teste.

Observações Finais

- No Relatório deve estar bem claro como foram organizados os experimentos de tal forma a realizar corretamente a avaliação dos modelos e a comparação entre os mesmos. Fornecer também uma descrição sucinta dos dados. No relatório mostrar os detalhes da obtenção dos hiper-parâmetros do modelo, se houver.
- Data de apresentação e entrega do projeto: **QUARTA-FEIRA 08/03/2023 a partir das 15 horas.**
- Colocar no google classroom o relatório do projeto, e os slides da apresentação.
- **NÃO COLOCAR ARQUIVO ZIPADO**
- Tempo de apresentação: **15 minutos** para cada equipe (rigoroso).
- **A presença de todos os membros de cada equipe é obrigatória durante a apresentação;**