
Projeto da disciplina

Prof. Leandro Almeida
lma3@cin.ufpe.br

Atividade

- ❖ Trabalho prático de mineração individual
- ❖ Executar todas as etapas do CRISP-DM utilizando scripts em python
- ❖ Construir relatório sobre as decisões tomadas a respeito das etapas. Porque das escolhas.
- ❖ Fazer experimentos com K-NN, Árvore de decisão simples, Random Forest, rede neural MLP e um Comitê de Redes Neurais e um comitê heterogêneo em python
 - ❖ Será necessário fazer buscas pela configuração dos comitês
 - ❖ Utilizar a validação cruzada com k-fold estratificado com $k=10$
 - ❖ Apresentar a acurácia média (e desvio padrão), matriz de confusão e os gráficos de caixa (boxplot). Analisar os resultados com relatório de classificação e apresentar testes de significância estatística
- ❖ Extrair uma árvore de decisão e identificar quais são os primeiros atributos utilizados na construção da árvore
 - ❖ Propor sugestões de decisões com base nos atributos encontrados
- ❖ Entregar relatório final contendo todas as decisões e discussões sobre cada etapa do CRISP-DM.

Bases de dados

- ❖ As seguintes bases de dados do (UCI Machine Learning) podem ser utilizadas:
 1. Census Income Data Set
 2. Bank Marketing Data Set
 3. default of credit card clients Data Set
 4. Letter Recognition Data Set