

MiniTeste-1

(para avaliação individual)

Para se submetido até as 23h59 do dia 18/11.

1. Considere o dataset CBrasil. Pretende-se elaborar um estudo para o desenvolvimento de um modelo de previsão para analisar as faltas às consultas nos vários centros de saúde do Rio de Janeiro. O objetivo é ter um estudo sobre possíveis modelos pra prever/identificar os pacientes que tem tendência a faltar às consultas marcadas. Temos de pré-processar os dados por forma a:
 - Eliminar atributos redundantes (ou com pouco valor informativo)
 - Forçar atributos booleanos ou categóricos a serem mesmo booleanos ou categóricos (e não interpretar como numéricos como estão nos dados!)

Depois de atingir um dataset com os elementos relevantes apresente resultados que permitam responder às seguintes questões:

- 1.1. Qual os atributos a considerar para construir o modelo de previsão?
Justifique.
- 1.2. Qual atributo com maior valor informativo?
2. Uso o WEKA e as suas implementações de NaiveBayes, BayesNet e J48 para responder às seguintes perguntas (apresente resultados obtidos por validação cruzada e variação de hiper-parâmetros dos três algoritmos):
 - 2.1. Qual o modelo que escolhia para implementar em tempo real dentro destes 3 (e suas variantes)? Justifique.
 - 2.2. Em termos de classe no-show=yes qual o melhor modelo? Justifique?
 - 2.3. Para o modelo derivado do algoritmo J48 mostra as várias árvores possíveis de obter por diferentes configurações de *pruning*. Tente explicar os vários desempenhos (dos vários modelos derivados).