Aula3 – Medidas orientadas à classe, Curvas ROC, etc

1) Investigue as consequências do uso do “aligeiramento de Laplace” no erro de um modelo de previsão J48 e na AUC para o dataset diabetes

Algoritmo J48 com Cross Validation:

RMSE: 0.4463

Positivos marcados como negativos: 108

Negativos marcados como positivos: 93

Area under de curve: 0.7514

Algoritmo J48 com Cross Validation + Laplace:

RMSE: 0.4337

Positivos marcados como negativos: 108

Negativos marcados como positivos: 93

Area under de curve: 0.7751

2) Compare naive bayes, redes bayesianas e J48 com o dataset diabetes em termos de precision, recall e F1.

Algoritmo NaiveBayes com Cross Validation:

RMSE: 0.4168

Positivos marked as negativos: 104 Negativos marked as positivos: 78

Precision mean: 0.759 Recall mean: 0.763

F1:

Algoritmo BayesNet com Cross Validation:

RMSE: 0.4208

Positivos marked as negativos: 105 Negativos marked as como positivos: 92

Precision mean: 0.741 Recall mean: 0.743

F1:

Algoritmo J48 com Cross Validation:

RMSE: 0.4463

Positivos marked as negativos: 108 Negativos marked as positivos: 93

Precision mean: 0.735 Recall mean: 0.738

F1:

3) Avalie o impacto dos métodos de pruning em J48 quando avaliado por resubstituição (treino) e por validação cruzada.

4) Qual o melhor classificador para a classe “tested\_negative”? E para a classe “tested\_positive”? Justifique.

5) Se o objetivo fosse identificar todos os diabéticos, qual deveria ser o modelo escolhido? Justifique.

6) Qual é o modelo com melhor qualidade de previsão para a classe “tested\_positive”? Justifique