Atividade 2

Experimenter

(a)

1) Percent_correct

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result-matrix "weka.experiment.ResultMat
Analysing:  Percent_correct
Datasets:   4
Resultsets: 6
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       05/05/2023 11:18


Dataset                   (1) rules.Ze | (2) trees (3) lazy. (4) trees (5) bayes (6) funct
----------------------------------------------------------------------------------------
pima_diabetes             (100)   65.11 |   74.49 v   73.86 v   76.10 v   75.75 v   77.10 v
iris-weka.filters.unsuper(100)   33.33 |   94.80 v   96.40 v   95.40 v   95.93 v   95.27 v
statlog-heart-dataset     (100)   55.56 |   77.52 v   80.26 v   82.15 v   83.81 v   83.30 v
breast-cancer             (100)   70.30 |   74.28     74.00 v   69.75     72.70     74.94 v
----------------------------------------------------------------------------------------
                                (v/ /*) |   (3/1/0)   (4/0/0)   (3/1/0)   (3/1/0)   (4/0/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) trees.J48 '-C 0.25 -M 2' -217733168393644444
(3) lazy.IBk '-K 5 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\
(4) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698
(5) bayes.NaiveBayes '' 5995231201785697655
(6) functions.SimpleLogistic '-I 0 -M 500 -H 50 -W 0.0' 7397710626304705059
```

2) AUC ROC

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -result-matrix "weka.experiment.ResultMa
Analysing:  Area_under_ROC
Datasets:   4
Resultsets: 6
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       05/05/2023 11:18


Dataset                   (1) rules.Z | (2) tree (3) lazy (4) tree (5) baye (6) func
-------------------------------------------------------------------------------------
pima_diabetes             (100)   0.50 |   0.75 v   0.77 v   0.83 v   0.82 v   0.83 v
iris-weka.filters.unsuper(100)   0.50 |   0.99 v   1.00 v   1.00 v   1.00 v   1.00 v
statlog-heart-dataset     (100)   0.50 |   0.79 v   0.86 v   0.90 v   0.90 v   0.90 v
breast-cancer             (100)   0.50 |   0.61 v   0.67 v   0.67 v   0.70 v   0.69 v
-------------------------------------------------------------------------------------
                              (v/ /*) |  (4/0/0)  (4/0/0)  (4/0/0)  (4/0/0)  (4/0/0)


Key:
(1) rules.ZeroR '' 48055541465867954
(2) trees.J48 '-C 0.25 -M 2' -217733168393644444
(3) lazy.IBk '-K 5 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last
(4) trees.RandomForest '-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1' 1116839470751428698
(5) bayes.NaiveBayes '' 5995231201785697655
(6) functions.SimpleLogistic '-I 0 -M 500 -H 50 -W 0.0' 7397710626304705059
```

Para essas bases, salvo o random forest para a base de câncer de mama, todos os resultados foram melhores a nível de acurácia e área sob a curva ROC. Dependendo da base, um algoritmo é ligeiramente melhor que os demais.

(b) Com base na AUC, todos são estatisticamente superiores que a baseline. Entretanto quando observamos em relação à acurácia, para a última base de dados, 3 algoritmos não

tem superioridade estatística em relação ao ZeroR.