

Assignment 2
(CSE 6363)
Nishant Chaudhary(1001568781)

1. What is the meaning of an "*" or a "v" in the results?

The annotation v or * indicates that a specific result is statistically better (v) or worse (*) than the baseline scheme (in this case, j48) at the significance level specified (currently 0.05).

2. What is the meaning of the (x/y/z) counters that appear at the bottom?

At the bottom of each column after the first column is a count (x/ y/ z) of the number of times that the scheme was better than (x), the same as (y), or worse than (z), the baseline scheme on the datasets used in the experiment.

3. Which algorithm performs significantly better than the base algorithm for the "german-credit" dataset?

Naïve Bayes performs better than the base algorithm for the “german-credit dataset” dataset.

4. Is there an algorithm that performs significantly worse than the base algorithm (J48)?

No

5. What can you say about the std deviations shown in the test?

The std deviation for the *contact-lenses* dataset was the highest so it shows that this data is the most spread out. The least spread out data is *german-credit* because the value of std deviation for this dataset is the least among the three. The dataset for *pima-dataset* is a bit more spread out than *german-credit* but pretty less as compared to contact-lenses.

The standard deviation of the attribute being evaluated can be generated by selecting the Show std. deviations check box and hitting Perform test again. The value (100) at the beginning of the dataset row represents the number of estimates that are used to calculate the standard deviation (the number of runs in this case).

Result:

Tester: weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -V -result-matrix "weka.experiment.ResultMatrixPlainText -mean-prec 2 -stddev-prec 2 -col-name-width 0 -row-name-width 25 -mean-width 2 -stddev-width 2 -sig-width 1 -count-width 5 -show-stddev -print-col-names -print-row-names -enum-col-names"

Analysing: Percent_correct

Datasets: 3

Resultsets: 3

Confidence: 0.05 (two tailed)

Sorted by: -

Date: 3/22/19, 7:41 PM

Dataset (2) trees.J48 '-C 0 | (1) rules.JRip ' (3) bayes.NaiveB

contact-lenses	(100)	83.50(22.54)	81.00(23.09)	76.17(25.54)
german_credit	(100)	71.25(3.17)	72.21(3.96)	75.16(3.48) v
pima_diabetes	(100)	74.49(5.27)	75.18(4.54)	75.75(5.32)

(v/ /*) | (0/3/0) (1/2/0)

Key:

(1) rules.JRip '-F 3 -N 2.0 -O 2 -S 1' -6589312996832147161

(2) trees.J48 '-C 0.25 -M 2' -217733168393644444

(3) bayes.NaiveBayes " 5995231201785697655