

CS795/895: Topics in Data Mining and Security

Summer 2009

Homework #1: Due: May 20, 2009 (Wednesday)

Using the Weka software, with the training data provided in vote.arff (in datasets-UCI), determine if the following person is a democrat or a republican based on the given voting record:

'y', 'y', 'y', 'y', 'n', 'y', 'y', 'y', 'n', 'y', 'n', 'n', 'n', 'y', 'n', ???

'y', 'y', 'y', 'y', 'n', 'n', 'y', 'y', 'y', 'y', 'y', 'n', 'y', 'n', 'y', 'y', ???

'n', 'y', 'n', 'y', 'y', 'y', 'n', 'n', 'n', 'y', 'n', 'y', 'y', 'y', 'n', 'y', ???

Use the following 25 methods. For each method, summarize the results (as given by Weka).

Finally, provide a conclusive summary as to which one (in your opinion) is the most accurate.

1. BayesNet
2. NaiveBayes
3. NaiveBayesSimple
4. NaiveBayesUpdatable
5. Multilayerperceptron
6. SimpleLogistic
7. VotedPerceptron
8. IB1
9. IBk
10. KStar
11. LWL
12. AdaBoostM1
13. Attribute selectedclassifier
14. Bagging
15. Grading
16. J48
17. ADTree
18. DecisionStump
19. LMT
20. NBtree
21. RandomForest
22. RandomTree
23. ConjunctiveRule
24. DecisionTable

25. ZeroR