Shelby Luttrell

Questions: Naïve Bayes model

(1) Programming: Use Naïve Bayes to predict labels of instances in the testing set (12 games)

based on the training set (24 games). Calculate Accuracy, Precision, Recall, and F1 score on the

testing result.

precision recall f1-score support

0 0.50 1.00 0.67 3

1 1.00 0.67 0.80 9

avg / total 0.88 0.75 0.77 12

(2) Write down prediction label of the 12 testing games as well as evaluation result in the PDF.

[1 0 1 0 1 0 0 1 1 0 1 0]

24 Win

25 Lose

26 Win

27 Lose

28 Win

29 Lose

30 Lose

31 Win

32 Win

33 Lose

34 Win

35 Lose

(3) Compare Naïve Bayes with ID3 and C4.5, which model is the best, which model performs the

worst? Can you explain why?

Bayes classifiers return probabilities, it is simpler to apply these results to a wide variety of tasks than if an arbitrary scale was used. It does not require large amounts of data before learning can begin. Naive Bayes classifiers are computationally faster when making decisions, so it performs the best in this case. ID3 and C4.5 require that the target feature will have only discrete values. As decision trees use the “divide and conquer” method, they tend to perform well if a few highly relevant characteristics exist, but less so if many complex connections are present. The ID3 algorithm uses a greedy search. It selects a test using the information gain criterion, and then never explores the possibility of alternate choices so it performs the worst. C4.5 can fill in missing values for information can be over fitted if the data is noisy.