

CSCE 5073 Data Mining --- Homework

Due Date --- N/A. No need to submit. Answers will be posted.

1. (10pts) Consider the following statistics collected from a training dataset with 100 tuples. The class label is rating with two domain values: good or bad

Gender rating	Married	number of tuples with good rating	number of tuples with bad
F	N	5	20
F	Y	0	20
M	N	15	10
M	Y	30	0

1) calculate the information gain when splitting on Gender (decision tree algorithm). Note that you can keep “log” in your results.

2) calculate the gini when splitting on Married.

3) Given a new test tuple with Gender = F and Married= N, predict the class label based on the naïve bayes classifier.