INFS4203/7203 Data Mining

Tutorial 1: Classification

1. For the following two examples, calculate the accuracy, precision and recall:

a. Assume a test dataset contains 1% catfish and 99% other types of fish. For a classifier that always makes prediction that a fish is not a catfish, what is the accuracy, precision and recall.

b. Calculate the precision, recall and F-measure for the following two binary classifiers M1 and M2, and discuss which one is better.

M1

●	Red
●	Not Red
●	Not Red
●	Red
●	Red
●	Red
●	Not Red
●	Red
●	Not Red
●	Not Red

M2

●	Red
●	Not Red
●	Not Red
●	Red
●	Not Red
●	Red
●	Not Red
●	Not Red
●	Not Red
●	Not Red

2. Given the training data as the following table and an X = ($age = youth,\ income = medium,\ student = yes,\ rating = fair$), predict the class label (yes/no) for X using the Naïve Bayesian Classification.

RID	AGE	INCOME	STUDENT	RATING	CLASS
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle-aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle-aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle-aged	Medium	No	Excellent	Yes
13	Middle-aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No