INFS4203/7203 Assignment 4

Semester 2, 2020

Marks:	6 marks
Due Data:	**23:59, 25 Oct 2020, Brisbane time**
What to submit:	a report in the **pdf** format
Where to submit:	Turnitin link via blackboard

There are two goals of this assignment:

1. Test your ability to model a real-world problem into a classification problem such that the techniques taught in class can be applied.
2. Compare the performance of different classification algorithms on a toy data.

The are 2 questions (Problem 1 and Problem 2) in this assignment. You must work on this assignment **individually**. The standard academic honesty rules apply. You are required to read and understand the School Statement on Misconduct, available on the School's website at:

http://ppl.app.uq.edu.au/content/3.60.04-student-integrity-and-misconduct

Task:

Problem 1 [2 marks] Find an application in your own discipline or major (e.g. IT, bioinformatics, commerce, engineering, …) where the classification methods can be applied. Briefly answer the following questions:

1.1 [1 mark] Briefly discuss the application. Explicitly describe the labels and the process of training data collection.

1.2 [0.5 mark] Briefly discuss how classification technique could benefit your application.

1.3 [0.5 mark] Briefly discuss ethical issues if applying the classification technique to the application?

Problem 2 [4 marks] The following **training data** (Table 1) shows whether the bank approved a credit card application based on the information of the applicant's job status (**nominal feature**), marital status (**nominal feature**) and annual income (**numerical feature**).

Table 1: Training set

Permanent Job	Marital Status	Annual Income	Approved?
Yes	Single	130K	Yes
No	Married	80K	No
No	Single	100K	Yes
Yes	Divorced	90K	Yes
No	Single	60K	No
Yes	Married	120K	Yes
Yes	Single	85K	Yes
No	Divorced	110K	No
Yes	Married	95K	Yes
No	Married	125K	Yes

Please answer the following 2 questions:

2.1 [3 marks] Construct the following two classifiers based on the training data to predict whether the bank will approve a credit card application given features "Permanent Job", "Marital Status" and "Annual Income".

a. Decision tree based on Gain Ratio
b. Naïve Bayes (using **Laplacian correction** if necessary)

Please **describe** the constructed decision tree in plain language, and use the constructed two classifiers (a and b) to **fill** the corresponding blanks in Table 2.

Table 2

Permanent Job	Marital Status	Annual Income	Prediction (Approved?)	
			Decision Tree	Naïve Bayes
No	Single	60K		
Yes	Married	100K		
Yes	Single	90K		
No	Divorced	95K		
No	Married	85K		

2.2 [1 mark] Using Table 3 as the test data, compare the **accuracy** and **F1** of the two constructed classifiers. Discuss briefly which classifier best fits the data.

Table 3

Permanent Job	Marital Status	Annual Income	Approved?
No	Single	60K	No
Yes	Married	100K	Yes
Yes	Single	90K	No
No	Divorced	95K	Yes
No	Married	85K	No