

CONFIDENTIAL



UNIVERSITI TUN HUSSEIN ONN MALAYSIA

**FINAL EXAMINATION
SEMESTER II
SESSION 2018/2019**

COURSE NAME : DATA MINING
COURSE CODE : BIT 33603
PROGRAMME CODE : BIT
EXAMINATION DATE : JUNE / JULY 2019
DURATION : 3 HOURS
INSTRUCTION : ANSWER ALL QUESTIONS

THIS QUESTION PAPER CONSISTS OF **FOUR (4)** PAGES

TERBUKA
CONFIDENTIAL

Q1 Define the following terms

- (a) Data mining (2 marks)
- (b) Data clustering (2 marks)
- (c) Pattern recognition (2 marks)
- (d) On-Line Analytical Processing (OLAP) (2 marks)

Q2 Describe the following data mining tasks

- (a) Classification and Prediction (4 marks)
- (b) Association Rule Mining (4 marks)
- (c) Sequential Pattern Discovery (4 marks)

Q3 Four classifiers are generated for the same training set, which has 100 instances. They have confusion matrices shown in **Table 2**.

Table 2: Confusion Matrices

Classifier A		Predicted class	
		+	-
Actual class	+	50	10
	-	10	30

Classifier B		Predicted class	
		+	-
Actual class	+	40	20
	-	1	39

Classifier C		Predicted class	
		+	-
Actual class	+	55	5
	-	5	35

Classifier D		Predicted class	
		+	-
Actual class	+	60	0
	-	20	20

- (a) Calculate the values of the True Positive rate and False Positive rate for each classifier. (8 marks)
- (b) Based on the answer in **Q5(a)**, identify which classifier would be the best? (2 marks)

TERBUKA

Q4 Calculate the support(s) and confidence(s) for each rules below which generated from market basket transaction in **Table 1**.

Table 1: Items and their Transactions

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

- (a) {Milk, Diaper} → {Beer} (3 marks)
- (b) {Diaper, Beer} → {Milk} (3 marks)
- (c) {Diaper} → {Milk, Beer} (3 marks)
- (d) {Milk} → {Diaper, Beer} (3 marks)

TERBUKA

Q5 Table 1 shows a dataset for making decision to Buy Computer.

Table 1: Buy Computer Dataset

ID	age	income	student	credit_rating	class: buy_computer
1	<=30	high	no	fair	no
2	<=30	high	no	good	no
3	31...40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	good	no
7	31...40	low	yes	good	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	good	yes
12	31...40	medium	no	good	yes
13	31...40	high	yes	fair	yes
14	>40	medium	no	good	no

- (a) Build a decision tree using Information Gain as the attribute selection measure. (20 marks)

- (b) Predict the class of the following new example using the decision tree in Q2(a):
 age<=30, income=medium, student=yes, credit_rating=fair. (6 marks)

- (c) Predict the class of the following new example using Naïve Bayes classification:
 age<=30, income=medium, student=yes, credit_rating=fair. (12 marks)

- END OF QUESTION -

TERBUKA