UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2016/2017

COURSE NAME	:	DATA MINING
COURSE CODE	:	BIT 33603
PROGRAMME CODE	:	BIT
EXAMINATION DATE	:	DECEMBER 2016 / JANUARY 2017
DURATION	:	2 HOURS AND 30 MINUTES
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

TERBUKA

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

Q1 State whether each of the following activities is a data mining task. Explain your answer.

(a) Dividing the customers of a company according to their profitability.
(2 marks)

(b) Computing the total sales of a company.
(2 marks)

(c) Sorting a student database based on student identification numbers.
(2 marks)

(d) Predicting the outcomes of tossing a (fair) pair of dice.
(2 marks)

(e) Predicting the future stock price of a company using historical records.
(2 marks)

Q2 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.`

(5 marks)

c) Predict the class of the following new example using Naïve Bayes classification:

`age<=30, income=medium, student=yes, credit_rating=fair.`

(10 marks)

Q3 Answer **Q3(a)** and **Q3(b)** based on the transactions in Table **2**.

Table 2: Concession Database

Transaction ID	Items
T1	hotdogs, buns, ketchup
T2	hotdogs, buns
T3	hotdogs, coke, chips
T4	chips, coke
T5	chips, ketchup
T6	hotdogs, coke, chips

(a) Find all frequent itemsets for minimal support 0.3.

(10 marks)

(b) Find all association rules using Apriori algorithm of minimal confidence 0.6. Sort the answers by confidence.

(20 marks)

Q4 Answer **Q4(a)** – **Q4(c)** based on the following points in a 10 by 10 feature space:

`A(2,10), B(2,5), C(8,4), D(5,8), E(7,5), F(6,4), G(1,2), H(4,9)`

(a) Calculate the distance matrix (in square root) based on Euclidean distance.

(5 marks)

(b) Suppose the initial seeds are A, D, and G. Show the new clusters based on the *k*-means algorithm for 1 epoch only.

(10 marks)

(c) Draw all the points and show the clusters after the first epoch.

(5 marks)

Q5 How data mining techniques can be used in Social Network Analysis? Give at least **TWO (2)** examples.

(5 marks)

- END OF QUESTIONS -