Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN IIUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2017/2018

COURSE NAME	:	DATA MINING
COURSE CODE	:	BIT 33603
PROGRAMME CODE	:	BIT
EXAMINATION DATE	:	DECEMBER 2017/JANUARY 2018
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

TERBUKA

THIS QUESTION PAPER CONSISTS OF **SIX (6)** PAGES

Q1 State whether or not each of the following activities is a data mining task. Explain your answer.

(a) Monitoring the heart rate of a patient for abnormalities.

(2 marks)

(b) Computing the total profit of health insurance industry.

(2 marks)

(c) Monitoring and predicting failures in a hydropower plant.

(2 marks)

(d) Dividing the customers of a company according to their salary.

(2 marks)

(e) Extracting the frequencies of a sound wave.

(2 marks)

Q2 (a) K-NN (k-nearest-neighbor) classifies are lazy classifiers. Explain what does this mean?

(2 marks)

TERBUKA

(b) Explain the goal of dimensionality reduction techniques?

(2 marks)

(c) Are the two clusters shown in **Figure Q2(c)** is well separated? Justify your answer.

(2 marks)

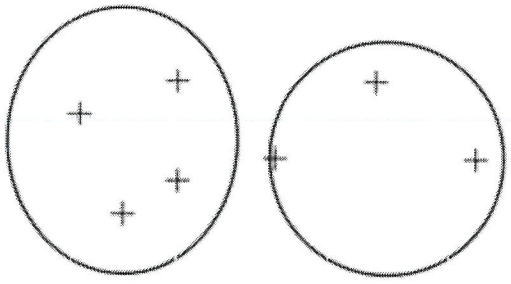

Figure Q2(c)

(d) A set of data is given as {3, 4, 2, 0, 21, 9, 1, 5}. Pre process the data into a new data set with [0, 0.9] range by using data normalization procedure as given in Equation 1.

$$D'(i) = \frac{D(i) - \min(D)}{\max(D) - \min(D)} * (upper - lower) + lower \quad \text{(Equation 1)}$$

(8 marks)

Q3 Four classifiers are generated for the same training set, which has 100 instances. They have confusion matrices shown in **Table 1**.

(a) Calculate the values of True Positive rate and False Positive rate for each classifier

(8 marks)

TERBUKA

(b) Calculate the value for Euclidean distance measure for each one.

(8 marks)

(c) Based on answer in **Q3 (b)**, identify which classifier would you consider the best?

(2 marks)

Table 1: Confusion Matrices

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

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

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

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

Q4 (a) Based on the following scenario in **Figure Q4(a),** draw and label a schematic diagram of a neural network architecture by considering the optimal number of weights for the network model is 48.

(10 marks)

Malaysian rubber industry is having a major problem in detecting symptoms of disease cause by fungus that destroys thousands hectares of Malaysia rubber plantings every year. Only 10,000 raw data had been collected manually in order to diagnose the symptoms whether infected, non-infected or neutral. Some major attributes that express the disease symptoms had also been identified as follows:

(1) The aggressiveness of fungus
(2) Size of fungus
(3) Type of fertilizer
(4) Humidity
(5) location

You have been chosen by Malaysia palm oil industry as data mining expert to solve their problem by using neural network classification task.

Figure Q4(a)

TERBUKA

(b) Based on **Figure Q4(b),** calculate the net output (before activation function) and neuron output (after activation function) for nodes 4, 5, and 6, by using the initial inputs and weights in **Table 2,** and initial bias in **Table 3.**

(18 marks)

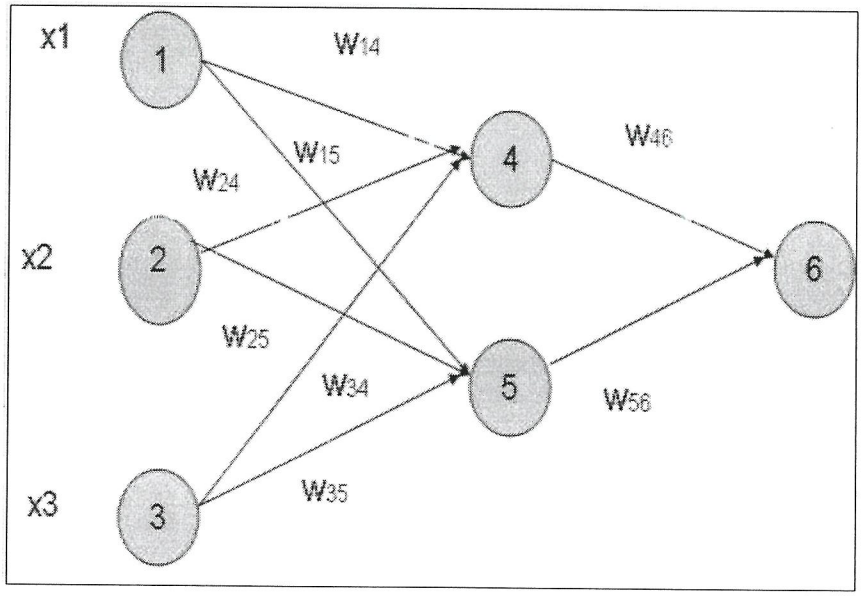

Figure Q4(b)

Table 2: Initial Inputs and Weights

x1	x2	x3	w14	w15	w24	w25	w34	w35	w46	w56
1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2

Table 3: Initial Bias

04	05	06
-0.4	0.2	0.1

TERBUKA

Q5 You are given a transaction data shown in the **Table 4** from a fast food restaurant. There are 9 distinct meal transactions (ID:1 – ID:9) and each transaction involves between 2 and 4 meal items.

Table 4: Meal Transactions

MEAL TRANSACTION ID	MEAL ITEMS
T1	M1, M2, M5
T2	M2, M4
T3	M2, M3
T4	M1, M2, M4
T5	M1, M3
T6	M2, M3
T7	M1, M3
T8	M1, M2, M3, M5
T9	M1, M2, M3

(a) Apply the Apriori algorithm to the transaction dataset and identify all frequent itemsets, with minimum support 2/9 (0.222). Show all of your work.

(20 marks)

(b) Find all association rules using Apriori algorithm of minimal confidence 7/9 (0.777).

(10 marks)

- END OF QUESTION -

TERBUKA