Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2021/ 2022

COURSE NAME	:	DATA MINING
COURSE CODE	:	BIT 33603
PROGRAMME CODE	:	BIT
EXAMINATION DATE	:	JULY 2022
DURATION	:	3 HOURS
INSTRUCTION	:	1. ANSWER ALL QUESTIONS. 2. THIS FINAL EXAMINATION IS AN **ONLINE ASSESSMENT** CONDUCTED VIA **CLOSED BOOK**. 3. STUDENTS ARE **PROHIBITED** TO CONSULT THEIR OWN MATERIALS OR ANY EXTERNAL RESOURCES DURING THE EXAMINATION CONDUCTED VIA CLOSED BOOK.

TERBUKA

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

Q1 (a) Differentiate between supervised and unsupervised learning?

(6 marks)

(b) State **TWO (2)** examples of technique in each of the mentioned learning.

(4 marks)

Q2 (a) Define the term k-means clustering.

(4 marks)

(b) Describe the algorithm to develop clusters.

(6 marks)

(c) Suppose that 5 objects and each object has location in x and y-axis as shown in Table 1. Group these objects into K=2 clusters based on their location.

Table 1: Location of the objects

	x-axis	y-axis
Object 1	1	1
Object 2	2	1
Object 3	4	3
Object 4	5	4
Object 5	3	6

(10 marks)

Q3 (a) Define artificial neural network.

(2 marks)

(b) Discuss over-fitting and outlier.

(4 marks)

(c) The 3-input neural network with the set of weights as $w_0 = 0.3$, $w_1 = -2.0$, $w_2 = 1.5$, and suppose that the input pattern is $x_0 = 1$, $x_1 = 1$, $x_2 = -1$, and the desired output is 1.

(i) Draw the structure diagram of neural network.

(5 marks)

(ii) Calculate the actual output.

(5 marks)

TERBUKA

(iii) Assuming that the weights are updated after each pattern and the value of η is **0.33**, Calculate the updated weights?

(8 marks)

(iv) Calculate new actual output.

(2 marks)

Q4 (a) Describe Association Rule.

(8 marks)

(b) Answer **Q4(b)(i)** and **Q4(b) (ii)** based on Table 2.

Table 2: Transaction on items {A,B,C,D,E}

Tid	Items
1	{A,B}
2	{A,B,C}
3	{B,C,D}
4	{B,C}
5	{A,B,C,D}
6	{B,D}
7	{B,E}
8	{B,D,E}

(i) Apply the Apriori Algorithm to compare the frequent item sets and with minimum support as 3.

(6 marks)

(ii) Write the algorithm to give all generators of closed frequent item sets and their closure.

(6 marks)

Q5 (a) Discuss information measure relate to decision tree structure.

(4 marks)

(b) Based on **Figure Q5(b)**, answer **Q5(b)(i) - Q5(b)(iii).**

A study on customer expenses is conducted and a dataset is given in Table 1. The study shows either customer will buy a house or not. The decision or the dependent variable is identified in the last column. Summary of all the entropy calculation are tabulated in Table 3,4,5 and 6.

Figure Q5(b)

TERBUKA

Table 3: Customer dataset

ID	Age	Income	Government Employee	Credit Rating	Buy House
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

Table 4: Entropy measure for root node

Attribute	Average Entropy
Age	0.6935
Income	0.9110
Government Employee	0.7885
Credit rating	0.8922

Table 5: Entropy measure for the branch node ((<=30)

Attribute	Average Entropy
Income	0.4000
Government Employee	0.0000
Credit rating	0.9510

Table 6: Entropy measure for the branch node (>40)

Attribute	Average Entropy
Income	0.951
Government Employee	0.951
Credit rating	0.000

(i) Construct a decision tree using the entropy measure.

(8 marks)

(ii) Convert the decision tree in **Q5(b)(i)** to a production rules.

(8 marks)

(iii) Determine the result of an old government employee with fair credit rating?

(4 marks)

- END OF QUESTIONS -

TERBUKA