



UNIVERSITI TUN HUSSEIN ONN MALAYSIA

**FINAL EXAMINATION
SEMESTER II
SESSION 2017/2018**

COURSE NAME : INTRODUCTION TO DATA MINING
COURSE CODE : BWB 43303
PROGRAMME CODE : BWQ
EXAMINATION DATE : JUNE / JULY 2018
DURATION : 3 HOURS
INSTRUCTION : ANSWER ALL QUESTIONS

THIS QUESTION PAPER CONSISTS OF **THREE (3)** PAGES

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

- (a) Dividing the customers of a company according to their profitability. (3 marks)
- (b) Predicting the outcomes of tossing a (fair) pair of dice. (3 marks)
- (c) Predicting the future stock price of a company using historical records. (3 marks)
- (d) Monitoring the heart rate of a patient for abnormalities. (3 marks)
- (e) Predicting the number of students that will enrol in Statistics Industry course for session 2018/2019 based on the offer letter. (3 marks)
- (f) Monitoring and predicting failures in a hydropower plant. (3 marks)
- (g) Grouping the students based on their CGPA. (3 marks)

- Q2**
- (a) Describe what data mining is. (2 marks)
 - (b) Based on your answer in **Q2(a)**, address the following:
 - (i) Is it another hype? (3 marks)
 - (ii) Is it a simple transformation of technology developed from databases, statistics, and machine learning? (3 marks)
 - (iii) Explain how the evolution of database technology led to data mining. (4 marks)
 - (iv) Distinguish the steps involved in data mining when viewed as a process of knowledge discovery. (14 marks)

TERBUKA

- Q3** (a) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are presented in **Table Q3(a)**. Answer the following.

Table Q3(a): Age

13	21
15	22
16	22
16	25
19	25
20	30
20	35

- (i) Transform the data using min-max normalization. The calculation need to be shown. (8 marks)
- (ii) Transform the data Z-score normalization. Show the process of transformation. (9 marks)
- (iii) Transform the data using decimal scaling. The calculation need to be shown. (7 marks)
- (iv) Comment on which method you would prefer to use and provide the reasons. (4 marks)
- Q4** (a) Given a decision tree, you have the option of (i) converting the decision tree to rules and then pruning the resulting rules, or (ii) pruning the decision tree and then converting the pruned tree to rules. What advantages does (i) have over (ii). (2 marks)
- (b) Discuss why tree pruning is useful in decision tree induction. (4 marks)
- (c) Elaborate the drawback of using a separate set of tuples to evaluate pruning. (4 marks)
- Q5** Differentiate the following approaches to clustering: partitioning methods, hierarchical methods, density-based methods, grid-based methods and model-based methods. Give example in each case. (15 marks)

– END OF QUESTIONS –