UNIVERSITY OF MALAYA

EXAMINATION FOR THE DEGREE OF MASTER OF DATA SCIENCE

ACADEMIC SESSION 2021/2022   : SEMESTER 2

WQD7005   : Data Mining

June 2022

Duration:From   18/06/2022,8.00 PM, to 25/06/2022, 11.00 PM for Question 1
                   18/06/2022, 8.00 PM to 18/06/2022, 11.00 PM for Question 2,3,4
                   and 5

INSTRUCTIONS TO CANDIDATES :

Answer **ALL** questions (50 marks).

(This question paper consists of 5 questions on 3 printed pages)

1 Individual Case Study

The clustering and association rule of data is an important data miming procedure. You may need to compare between associaton rule vs clustering algorithms. Choose any dataset (except census2000 and DUNGAREE) and present the comparative analysis between clustering algorithms vs association rules, keeping in mind all the stages of SEMMA by using the concepts of data clustering or asscoiation rules studied in this course. The emphasis is on the use of data mining methodology in real world applications.

(25 marks)

Requirements and grading:

Each student must submit one report for question 1. The report should have the following structure,

1. Title

2. Abstract

3. Introduction

4. Review of relevant literature

5. Description of solution

6. Results and discussion

7. Conclusion

8. Bibliography

2 Suppose that a data warehouse, **University Hospital** consists of four dimensions **time**, **doctor, ward** and **patient** and along with two measures: **charge** and **count**.

Draw a schema diagram for the above data warehouse using Star schema.

(10 marks)

3 Define your words of Data Mining*.*

*(*4 marks*)*

4 Discuss the differences between database processing query and data mining processing query with own three examples.

(6 marks)

5   Data quality can be assessed in terms of several issues, including accuracy, completeness, and consistency. Propose two other dimensions of data quality. Discuss how data quality assessment that you address these 5 issues in your group project dataset.

(5 marks)

END