

UNIVERSITY OF MALAYA

EXAMINATION FOR THE DEGREE OF MASTER OF DATA SCIENCE

ACADEMIC SESSION 2022/2023 : SEMESTER 1

WQD7005 : Data Mining

Duration: From 14/01/2023, 8.00 PM, to 28/01/2023, 11.00 PM

INSTRUCTIONS TO CANDIDATES :

Answer **One** question (25 marks).

(This question paper consists of 1 question on 2 printed pages)

1 Individual Case Study

Association analysis is primarily frequency analysis performed on a transaction dataset. Since datasets for most practical problems are large you need clever algorithms like *Apriori* to manage association analysis.

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data analysis in large dataset.

You may need to compare association analysis vs clustering. Choose any dataset (except census2000 and DUNGAREE) and present the comparative analysis between clustering and association analysis, keeping in mind all the stages of SEMMA by using the concepts of clustering and association analysis 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

END