UNIVERSITY OF MALAYA

MIDTERM EXAMINATION FOR THE DEGREE OF MASTER OF DATA SCIENCE

ACADEMIC SESSION 2019/2020 : SEMESTER II

WQD7005 : Data Mining

15 May 2020 TIME : 3 Hours

INSTRUCTIONS TO CANDIDATES :

Answer **ALL** questions (100 marks).

(This question paper consists of 6 questions on 3 printed pages)

Instructions: Work individually, submission via Spectrum.

. 1. You are required to write code that will crawl the WWW (your familiar domain) of a particular website and collect data from social media or from report or news and so on and to produce dataset.

(20 marks)

1. After that you are required to write code to modify and/or select your the attributes to perform the following tasks:

* Data Cleaning
* Data Integration
* Data Transformation
* Data Reduction

Hint : Make sure that use graph to help you select appropriate attributes to perform the above tasks.

(20 marks)

1. You are required to write data cleaning code using the appropriate attribute (Question 2). In order to achieve the task, you are going to cover the following steps:

* Importing required libraries
* Loading Data
* Before data cleaning, plot your result (attribute)
* Data cleaning
* After data cleaning, plot your result (attribute)

(10 marks)

1. You are required to write data integration code using the appropriate attribute (Question 2). In order to achieve the task, you are going to cover the following steps:

* Importing required libraries
* Loading Data
* Before data integration, plot your result (attributes from www, social media, report, news)
* Data integration
* After data integration, plot your result (attributes from www, social media, report, news)

(10 marks)

1. You are required to write data transformation code using the appropriate attribute (Question 2). In order to achieve the task, you are going to cover the following steps:

* Importing required libraries
* Loading Data
* Before data transformation, plot your result (attributes)
* Data transformation
* After data transformation, plot your result (attributes)

Hint : Normalization, Aggregation and Generalization

(20 marks)

1. You are required to write data reduction code using the appropriate attribute (Question 2). In order to achieve the task, you are going to cover the following steps:

* Importing required libraries
* Loading Data
* Before data reduction, plot your result (attributes)
* Data reduction
* After data reduction, plot your result (attributes)

Hint : Dimensionality Reduction

(20 marks)

Submissions:

The student is expected to submit answers to each question individually, and submit the document in PDF format. The student can include online materials, screenshots, and codes to support your answer.

**END**