B.Sc. TE 1 Yr. 1st Semester
B.Sc. TE 2 Yr. 3rd Semester

07 March 2019 (Afternoon)

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

MID SEMESTER EXAMINATION **WINTER SEMESTER, 2018-2019**
DURATION: 1 Hour 30 Minutes **FULL MARKS: 75**

CSE 4775: Introduction to Data Mining

Programmable calculators are not allowed. Do not write anything on the question paper.

There are **4 (four)** questions. Answer any **3 (three)** of them.

Figures in the right margin indicate marks.

1. a) What is data mining? Describe the steps involved in data mining when viewed as a process of knowledge discovery. 15

   b) Describe three challenges to data mining regarding *"Scalability & Efficiency"* and *"Data Mining & Society"*. 10

2. a) Given two objects represented by the tuples (-2, 1, 42, 10) and (21, 0, -6, 10): 5×4
   - i. Compute the Euclidean distance between the two objects.
   - ii. Compute the Manhattan distance between the two objects.
   - iii. Compute the Minkowski distance between the two objects, using $h = 4$.
   - iv. Compute the supremum distance between the two objects.
   - v. Which distance among them is the most suitable one. Justify your Answer.

   b) What are the different types of data used in Data Mining applications? 5

3. a) Briefly outline how to compute the dissimilarity between objects described by mixed attribute. 12

   b) What is *Interquartile Range*? How IQR is used for outlier analysis? 7

   c) What are the conditions a pattern should fulfill to be interesting? 6

4. a) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order): 18

   13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

   - i. What is the *mean* and *Median* of the data?
   - ii. What is the *mode* of the data? Comment on the data's modality.
   - iii. What is the midrange of the data?
   - iv. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
   - v. Give the five-number summary of the data.
   - vi. Show a boxplot of the data.

   b) Differentiate between *Data Matrix* and *Dissimilarity Matrix* with appropriate example. 7