UECS3213 / UECS3453 DATA MINING

SESSION: January 2019

TUTORIAL 1

Chapter 1: Introduction to Data Mining

1. Define data mining.

2. Describe the steps involved in data mining when viewed as a process of *knowledge discovery*.

3. Discuss whether or not each of the following activities is a *data mining task*.

   a) Dividing the customers of a company according to their gender.
   b) Dividing the customers of a company according to their profitability.
   c) Computing the total sales of a company.
   d) Sorting a student database based on student identification numbers.
   e) Predicting the outcomes of tossing a (fair) pair of dice.
   f) Predicting the future stock price of a company using historical records.
   g) Monitoring the heart rate of a patient for abnormalities.
   h) Monitoring seismic waves for earthquake activities.
   i) Extracting the frequencies of a sound wave.

4. Suppose that you are employed as a data mining consultant for an Internet search engine company. Describe how data mining can help the company by giving specific examples of how techniques, such as *clustering*, *classification*, *association rule mining*, and *anomaly detection* can be applied.

5. Define each of the following data mining functionalities: *characterization, discrimination, association and correlation analysis, classification, prediction, clustering*, and *evolution analysis*.

   Give examples of each data mining functionality, using a real-life database that you are familiar with.

Discussion Questions

6. What is the *difference* between discrimination and classification? Between characterization and clustering? Between classification and prediction? For each of these pairs of tasks, how are they similar?

7. What are the major *challenges* of mining a huge amount of data (such as billions of tuples) in comparison with mining a small amount of data (such as a few hundred tuple data set)?

The End