

## MODEL QUESTION PAPER

Reg. No:.....

Name:.....

### **Second Semester M.Tech. Degree Examination, 2014** **Branch: COMPUTER SCIENCE AND ENGINEERING** **RCD 2001: DATA WAREHOUSING & MINING**

Time: 3 Hours

Max. Marks: 60

*Instructions: Answer **two** questions from each module. All questions carry **equal** marks.*

#### **Module 1**

- I.
  - a. Define each of the following *data mining functionalities*: characterization, discrimination, association and correlation analysis, classification, prediction, clustering, and evolution analysis.
  - b. Describe three challenges to data mining regarding *data mining methodology and user interaction issues*.
  
- II.
  - a) Suppose that a data warehouse consists of the three dimensions *time*, *doctor*, and *patient*, and the two measures *count* and *charge*, where *charge* is the fee that a doctor charges a patient for a visit.
    - i. Enumerate three classes of schemas that are popularly used for modeling data warehouses.
    - ii. Draw a schema diagram for the above data warehouse using one of the Schema classes listed in (a).
    - iii. Starting with the base cuboid [*day*, *doctor*, *patient*], what specific *OLAP operations* should be performed in order to list the total fee collected by each doctor in 2004?

- iv To obtain the same list, write an SQL query assuming the data are stored in a relational database with the schema *fee (day, month, year, doctor, hospital, patient, count, charge)*.
- b) In data warehouse technology, a multiple dimensional view can be implemented by a relational database technique (*ROLAP*), or by a multidimensional database technique (*MOLAP*), or by a hybrid database technique (*HOLAP*). Briefly describe each implementation technique.

**III.**

(i) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):

- (a) Compute the *Euclidean distance* between the two objects.
- (b) Compute the *Manhattan distance* between the two objects.
- (c) Compute the *Minkowski distance* between the two objects, using  $q = 3$ .

(ii) Differentiate between supervised and unsupervised learning in neural networks. Explain how neural network can be used in data mining.

**Module 2**

**IV.**

- a. Write short notes on Hierarchical clustering algorithms.
- b. Discuss in detail squared Error clustering algorithm.

**V.**

- a. Explain *k-means* and *k-medoids* algorithms that perform effective clustering. Illustrate the strength and weakness of *k-means* in comparison with the *k-medoids* algorithm.
- b. For the given sequence database, generate candidate set using SPADE technique.

| Sequence_id | Sequence          |
|-------------|-------------------|
| 1           | (a(abc)(ac)d(cf)) |
| 2           | ((ad)c(bc)(ae))   |
| 3           | ((ef)(ab)(df)cb)  |
| 4           | (eg(af)cbc)       |

**VI.**

- a. Represent a decision tree for a student record database
- b. What is Decision tree? Elaborate on classification done using Decision tree based algorithms.

### **Module 3**

**VII.**

- c. Give some examples of spatial queries.
- d. What is temporal mining and how is it different from spatial mining.

**VIII.** a. Explain Harvest System

- b. Name some algorithms used for search engine.

**IX.** A database has five transactions. Let  $min\ sup = 60\%$  and  $min\ con\ f = 80\%$ .

*TID items bought*

T100 { M, O, N, K, E, Y }

T200 { D, O, N, K, E, Y }

T300 { M, A, K, E }

T400 { M, U, C, K, Y }

T500 { C, O, O, K, I, E }

Find all frequent itemsets using Apriori and FP-growth, respectively.  
Compare the efficiency of the two mining processes.