

GUJARAT TECHNOLOGICAL UNIVERSITY
M. E. - SEMESTER – II • EXAMINATION – SUMMER • 2013

Subject code: 1720108**Date: 05-06-2013****Subject Name: Data Mining and Data Warehousing****Time: 10.30 am – 01.00 pm****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1** (a) What is Data Mining? Write down short note on KDD process. **07**
 (b) List and describe five primitives for specifying data mining task. **07**

- Q.2** (a) Describe the difference between the following approaches for the integration of a data mining system with a database or data warehouse system: no coupling, loose coupling, semi tight coupling and tight coupling. State which approach you think is most popular and why? **07**
 (b) What do you mean by text mining? Explain various issues involved in it. **07**

OR

- (b) A group of 12 sales price records has been sorted as follows: **07**
 5,10,11,13,15,35,50,55,72,92,204,215
 Partition them into three bins by each of following methods.
 1. Equal Frequency (Equi-depth) Partitioning
 2. Equal Width Partitioning
 3. Clustering
- Q.3** (a) Data warehouse consists of the four dimensions, date, spectator, location and game and the two measures, count and change, where change is the fare that a spectator pays when watching a game on a give date. Spectators may be students, adults, or seniors, with each category having its own charge rate. **07**
 1) Draw a Star Schema diagram for the data warehouse.
 (b) Explain K-means Algorithm with its limitation. **07**

OR

- Q.3** (a) A base cuboid has three dimensions A, B, C with the following number of cells: $|A| = 1,000,000$, $|B|=100$ and $|C|=1000$. Suppose that each dimension is evenly partitioned into 10 portions for chunking. **07**
 1) Assuming each dimension has only one level, draw the complete lattice of the cube.
 2) If each cube cell stores one measure with 4 bytes, what is the total size of the computed cube if the cube is dense?
 (b) The Apriori algorithm makes use of prior knowledge of subset support properties. **07**
 1) Prove that all non-empty subsets of a frequent itemset must also be frequent.
 2) Prove that the support of any non-empty subset s' of itemset s must be at least as great as the support of s .

- Q.4** (a) Discuss why analytical data characterization is needed and how it can be performed. Compare the result of two induction methods. **07**
 1) With relevance Analysis
 2) Without relevance Analysis
 (b) Write down short note on all search strategies for mining multiple-level associations with reduced support. **07**

OR

- Q.4 (a)** Why naïve Bayesian classification is called ñnaïveö? Briefly outline the major idea of naïve Bayesian classification. **07**
- Q.4 (b)** What is Market Basket Analysis? Explain Association Rules with Confidence and support. **07**
- Q.5 (a)** Briefly outline Clustering Using Representatives (CURE) with its merits and demerits. **07**
- (b)** Compare the merits and demerits of Eager Classification Versus Lazy Classification. **07**

OR

- Q.5 (a)** What is wave cluster? Explain Clustering using Wavelet transformation. **07**
- (b)** Briefly explain Attribute Oriented Induction approach with an example. **07**
