Seat No.: _____ **Enrolment No._____**

GUJARAT TECHNOLOGICAL UNIVERSITY

ME Semester –II Examination Dec. - 2011

Subject code: 1720108 **Date: 16/12/2011**
Subject Name: Data Mining & Data Warehousing
Time: 02.30 pm – 05.00 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)**	1. Define: data mining, data warehousing, data mart.	**03**
		2. What is difference between OLTP and OLAP?	**04**
	(b)	Explain various data repositories on which data mining techniques are applied.	**07**
Q.2	**(a)**	Explain various data cleaning methods in brief.	**07**
	(b)	Suppose a group of 12 sales price records has been sorted as follows. 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215 Partition them into three bins by each of the following methods. (i) equal frequency (ii) equal width (iii) clustering	**07**
		OR	
	(b)	What do you mean by numerosity reduction? Explain various methods to achieve it.	**07**
Q.3	**(a)**	1. Define: fact table, cuboid.	**02**
		2. What do you mean by concept hierarchy? Show its application with suitable example.	**05**
	(b)	Suppose a data warehouse for GTU consists of following four dimensions: student, course, semester, faculty with two measures: count, avg_grade. (i) Draw a snowflake schema for data warehouse. (ii) Starting with base cuboid (student, course, semester, faculty), what specific OLAP operations should one perform in order to list average grade of CSE course for each GTU student.	**07**
		OR	
Q.3	**(a)**	What do you mean by frequent item-set mining for market basket analysis? Explain apriori algorithm for the same with suitable example.	**07**
	(b)	What do you mean by closed frequent item set? What is its application? Which are various searching methods for it?	**07**
Q.4	**(a)**	Define: correlation. Explain any three methods to find correlation among items using suitable example.	**07**
	(b)	Compare association and classification. Briefly explain associative classification with suitable example.	**07**
		OR	
Q.4	**(a)**	Compare various attribute selection measures for decision tree with suitable example.	**07**

Q.4	**(b)**	What is prediction? Explain various regression methods for it.	**07**
Q.5	**(a)**	Explain cross validation and bootstrap methods for evaluating accuracy of classifier/predictor.	**07**
	(b)	Differentiate between partitioning and hierarchical methods for clustering. Explain any one hierarchical method in brief with suitable example.	**07**

OR

Q.5	**(a)**	Illustrate strength and weakness of k-mean in comparison with k-medoid algorithm.	**07**
	(b)	Define following terms w.r.t. text mining: precision, recall, document ranking, stop list, term-frequency matrix, inverse document frequency and tokenization.	**07**

****