GUJARAT TECHNOLOGICAL UNIVERSITY
ME - SEMESTER– II (Old course)• REMEDIAL EXAMINATION – SUMMER 2015

Subject Code: 1720108   **Date:15/05/2015**

Subject Name: Data Mining & Data Warehousing
Time: 02:30 pm to 5:00 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)**	In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.	**07**
	(b)	(I) How is data warehouse different from a database? How are they similar?	**04**
		(II) Discuss issues to consider during data integration.	**03**
Q.2	**(a)**	What are the differences between the three main types of data warehouse usage: information processing, analytical processing, and data mining? Give the differences between OLTP and OLAP	**07**
	(b)	List and describe the five primitives for specifying a data mining task	**07**

OR

	(b)	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?	**07**
Q.3	**(a)**	What do you mean by classification? Explain classification by Decision tree induction?	**07**
	(b)	(I) Define Predictive model.	**02**
		(II) Define descriptive model	**02**
		(III) Explain the term summarization	**03**

OR

Q.3	**(a)**	(I) Discuss the importance of similarity metric clustering? Why is it difficult to handle categorical data for clustering?	**04**
		(II) Define Binary variables? And what are the two types of binary variables?	**03**
	(b)	Explain the following:	**07**

           (i) Gini Index
           (ii) Entropy
           (iii) Gain
           (iv) Gain Ratio

Q.4	**(a)**	Association rule mining often generates a large number of rules. Discus effective methods those can be reduce the number of rules generated while still preserving most of the interesting rules.	**07**
	(b)	What do you mean by Cluster Analysis? Explain K-Medoids Partition method in detail.	**07**

OR

Q.4	**(a)**	Explain Apriori Algorithms with an example.	**07**
	(b)	What is the difference between õsupervisedö and unsupervisedö learning scheme? Explain DBSCAN in detail.	**07**
Q.5	**(a)**	What do you mean by text mining? Explain different approaches for text mining based on the kinds of data they take as input irregular.	**07**
	(b)	Describe Density-Based Local Outlier Detection.	**07**

OR

Q.5	**(a)**	Define following terms w.r.t. text mining: precision, recall, document ranking, stop list, term-frequency matrix, inverse document frequency and tokenization.	**07**
	(b)	Explain Agglomerative and Divisive Hierarchical Clustering.	**07**