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

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VII (OLD) - EXAMINATION – SUMMER 2017

Subject Code: 171601 **Date: 09/05/2017**

Subject Name: Data warehousing and Data Mining

Time: 02:30 PM to 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)**	Explain what is meant by 'multi-dimensional data model' for data warehouse. Discuss Star and Snowflake schemas using suitable examples.	**07**
	(b)	What is KDD? With the help of a neat diagram explain the steps in the KDD process.	**07**
Q.2	**(a)**	Explain the role of 'metadata repository' in a data warehouse. Discuss the contents of the metadata repository.	**07**
	(b)	What are the reasons for presence of 'noise' in data collected for mining? Explain the methods to deal with noise.	**07**
		OR	
	(b)	Explain the need for Data Reduction during pre-processing and discuss various strategies for data reduction.	**07**
Q.3	**(a)**	Explain the steps of the Apriori algorithm for mining frequent itemsets with candidate generation.	**07**
	(b)	What is 'data discretization'? Explain how a continuous numeric attribute can be discretized.	**07**
		OR	
Q.3	**(a)**	Explain various methods to improve the efficiency of the Apriori- based mining.	**07**
	(b)	What is an 'outlier'? How do outliers impact the results of mining? Explain any one method to detect outliers.	**07**
Q.4	**(a)**	Explain the following attribute selection measures for decision trees: (i) Information Gain (ii) Gain Ratio.	**07**
	(b)	Explain the working of the 'k-means' clustering algorithm. Also state its limitations.	**07**
		OR	
Q.4	**(a)**	Write a short note on 'Linear Regression'.	**07**
	(b)	Discuss density-based clustering in brief. Name any two algorithms for density-based clustering.	**07**
Q.5	**(a)**	Discuss the challenges in mining the World Wide Web, and explain the role of search engines.	**07**
	(b)	Name and explain types of OLAP servers.	**07**
		OR	
Q.5	**(a)**	Explain how trend analysis can be performed by mining Time-Series data.	**07**
	(b)	Discuss the characteristics and limitations of Neural Networks. Mention any two suitable applications of neural networks.	**07**

****