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. (a) Explain what is meant by 'multi-dimensional data model' for data warehouse. 0.1 07 Discuss Star and Snowflake schemas using suitable examples. What is KDD? With the help of a neat diagram explain the steps in the KDD 07 process. Explain the role of 'metadata repository' in a data warehouse. Discuss the **Q.2** 07 contents of the metadata repository. (b) What are the reasons for presence of 'noise' in data collected for mining? **07** Explain the methods to deal with noise. OR (b) Explain the need for Data Reduction during pre-processing and discuss various 07 strategies for data reduction. 0.3 (a) Explain the steps of the Apriori algorithm for mining frequent itemsets with 07 candidate generation. What is 'data discretization'? Explain how a continuous numeric attribute can **07** be discretized. OR (a) Explain various methods to improve the efficiency of the Apriori- based 0.3 07 (b) What is an 'outlier'? How do outliers impact the results of mining? Explain any 07 one method to detect outliers. 0.4 (a) Explain the following attribute selection measures for decision trees: **07** (i) Information Gain (ii) Gain Ratio. (b) Explain the working of the 'k-means' clustering algorithm. Also state its **07** limitations. OR (a) Write a short note on 'Linear Regression'. 0.4 07 (b) Discuss density-based clustering in brief. Name any two algorithms for density-**07** based clustering. Discuss the challenges in mining the World Wide Web, and explain the role of 0.5 (a) 07 search engines. **(b)** Name and explain types of OLAP servers. **07** OR (a) Explain how trend analysis can be performed by mining Time-Series data. **Q.5** 07 (b) Discuss the characteristics and limitations of Neural Networks. Mention any 07 two suitable applications of neural networks.

\*\*\*\*\*\*