

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER-VII • EXAMINATION – SUMMER • 2014

Subject Code: 171601

Date: 22-05-0214

Subject Name: Data Warehousing and Data Mining

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) What is Data Mining? List Challenges to data mining regarding data mining methodology and user- interaction issues. **07**
(b) Write short note on Spatial, Legacy and Multimedia Database. **07**
- Q.2** (a) Differentiate between Operational Database System and Data Warehouse. **07**
(b) What is Measures? List and explain types of measures. **07**
OR
(b) Define following terms: **07**
1. Data Mart
2. Enterprise Warehouse
3. Virtual Warehouse
- Q.3** (a) List and describe methods for handling missing values in data cleaning. **07**
(b) What is Concept Hierarchy? List and explain types of Concept Hierarchy. **07**
OR
Q.3 (a) What is Noisy data? List and describe data smoothing techniques. **07**
(b) Why naïve Bayesian classification is called “naïve”? Briefly outline the major ideas of naïve Bayesian classification. **07**
- Q.4** (a) Explain with an example attribute removal and attribute generalization. **07**
(b) State how the partitioning method may improve the efficiency of association mining. **07**
OR
Q.4 (a) Discuss why analytical data characterization is needed and how it can be performed. **07**
(b) Briefly outline the major steps of decision tree classification. **07**
- Q.5** (a) What is Cluster Analysis? List and explain requirements of clustering in data mining. **07**
(b) Explain different types of web mining with suitable example. **07**
OR
Q.5 (a) Write the steps of the k-medoids clustering algorithm with limitation. **07**
(b) Explain the information retrieval methods used in text mining. **07**
