

**GUJARAT TECHNOLOGICAL UNIVERSITY**  
**MCA - SEMESTER- IV • EXAMINATION – SUMMER 2017**

**Subject Code: 2640005****Date: 08/06/2017****Subject Name: Data Warehousing and Data Mining****Time: 10.30 AM TO 01.00 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) Do as directed
- (i) Define Multi feature Cube. 02
- (ii) Explain 4 Key words of Data Warehouse. 02
- (iii) Explain in brief Star, Snowflake and Galaxy schema with example. 03
- (b) Explain KDD process with its steps. 07
- Q.2** (a) What is the need for outlier analysis? Discuss various methods of Outlier detection 07
- (b) Discuss data cleaning techniques for missing values and Noisy data. 07
- OR**
- (b) Explain in detail Dimensionality reduction and Numerosity reduction. 07
- Q.3** (a) What is Apriori Property? Explain Apriori algorithm. 07
- (b) Explain the following:
- i) Concept Hierarchy 04
- ii) Self-exp, in-Exp, Path-exp 03
- OR**
- Q.3** (a) Discuss various techniques used for Improving the Efficiency of Apriori. 07
- (b) Explain the following:
- i) Attribute Removal and Attribute Generalization 04
- ii) Data Mart 03
- Q.4** (a) What is Data mining? Discuss major issues in data mining. 07
- (b) Explain in brief Boosting and Bagging. 07
- OR**
- Q.4** (a) Explain Decision tree algorithm with terms “Tree Induction” and “Tree Pruning”. 07
- (b) Discuss Holdout, random subsampling, cross validating and the bootstrap techniques for evaluating the Accuracy of a Classifier or Predictor. 07
- Q.5** (a) Explain k-mean algorithm for clustering. 05
- (b) (I) Describe application of Data Mining in Financial Data Analysis. 05
- (II) Discuss OLAP Servers. 04
- OR**
- Q.5** (a) Explain k-medoid algorithm for clustering. 05
- (b) (I) Describe application of Data Mining in Intrusion detection. 05
- (II) Differentiate Between OLAP and OLTP. 04

\*\*\*\*\*