Seat No.: _____   Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA - SEMESTER– IV • EXAMINATION – WINTER 2016

Subject Code: 2640005   **Date:26/10/ 2016**
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)**	Explain the following terms: 1. Data Mart? 2. Snowflake Schema? 3. Base Cuboid 4. Galaxy Schema 5. Smoothing 6. Decision tree pruning 7. OLTP	**07**
	(b)	Explain 3-Tire Data warehouse architecture with complete diagram.	**07**
Q.2	**(a)**	Explain Data Cube with all Operation in brief.	**07**
	(b)	Describe the term "Data Mining". Explain the process of Knowledge discovery from database.	**07**
		OR	
	(b)	Explain Data Discretization and Concept Hierarchy generation in brief.	**07**
Q.3	**(a)**	Explain following terms: 1. Binning Method 2. How can you handle missing values in data set?	**07**
	(b)	Suppose that a data warehouse consists of the four dimensions, date, spectator, location, and game, and the two measures, count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate. 1. Draw a star schema diagram for the data warehouse. 2. Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at SP Stadium in 2014? 3. Bitmap indexing is useful in data warehousing. Taking this cube as an example, briefly discuss advantages and problems of using a bitmap index structure.	**07**
		OR	
Q.3	**(a)**	What is classification? Also explain Supervised and Unsupervised learning in detail	**07**
	(b)	Describe Decision Tree Induction algorithm. You can describe it with the help of an example. How are the Rules induced from the Decision Tree?	**07**
Q.4	**(a)**	How can you generalize description resulting from Attribute Oriented Induction?	**07**
	(b)	Discuss two Ensemble methods for increasing the accuracy of classifier.	**07**
		OR	
Q.4	**(a)**	What is the purpose of Apriori algorithm? Explain with suitable example.	**07**
	(b)	What is Cluster analysis? Explain different types of cluster analysis.	**07**

1

Q.5 **(a)** Explain following attribute selection measure. **07**
1. Information Gain
2. Gini Index

(b) i. Differentiate Star Schema and Snow-Flake Schema **03**
ii. List and explain different types of OLAP servers. **04**

OR

Q.5 **(a)** Discuss the application of Data Mining for Telecommunication Industry **07**
(b) What are the techniques used in data reduction? Explain one of them in detail. **07**

****