Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA - SEMESTER–VI • EXAMINATION – SUMMER 2013

Subject Code: 640005 **Date: 31-05-2013**

Subject Name: Data Warehousing and Data Mining (DWDM)

Time: 10.30 am - 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) Answer the followings: **07**

1) What are the steps in the data mining process?
2) Define: Fact table
3) Define: Dimensional table
4) What is descriptive and predictive data mining?
5) What is decision tree pruning?
6) Define the merits of data warehouse.
7) State the disadvantages of K-means algorithm.

(b) Answer the followings: **07**

1) A data warehouse is said to contain a 'time-varying' collection of data because _____.
   (A) Its contents vary automatically with time
   (B) Its life-span is very limited
   (C) It contains historical data
   (D) Its content has explicit time-stamp
2) Dimension table are also known as _____.
   (A) Pivot table
   (B) Lookup table
   (C) Cross-tab
   (D) Fact Table
3) _____ is not an OLAP operation.
   (A) Slice and dice
   (B) Bitmap indexing
   (C) Pivoting
   (D) Roll up and drill down
4) Concept / class description is one of the _____of data mining.
   (A) Techniques
   (B) Needs
   (C) Functionalities
   (D) Issues
5) _____ is a type of histogram.
   (A) V-optimal
   (B) Min-Max
   (C) Equi-bucket
   (D) None of these

6) Which algorithm mines frequent itemsets without candidate generation?
   (A) Apriori
   (B) FP tree
   (C) FP growth
   (D) None of these
7) The ID3 algorithm stands for _____
   (A) Iteration Dichotomiser
   (B) Iterative Dichotomiser
   (C) Initial Dichotomiser
   (D) Inefficient Dichotomiser

Q.2	**(a)**	Give the comparison between OLTP and OLAP systems.	**07**
	(b)	Draw and explain the different types of multidimensional schemas with suitable example.	**07**
		OR	
	(b)	Describe and explain the measures with their categorization and computation in data warehouse.	**07**
Q.3	**(a)**	Discuss the various ways of handling missing values during data cleaning.	**07**
	(b)	Explain data reduction in data mining.	**07**
		OR	
Q.3	**(a)**	What is noise? Describe the possible reasons for noisy data. Explain the different techniques to remove the noise from data.	**07**
	(b)	List and describe major issues in data mining.	**07**
Q.4	**(a)**	Write and explain Apriori algorithm.	**07**
	(b)	Discuss the following as attribute selection measure: (1) Information gain (2) Gain ratio (3) Gini index	**07**
		OR	
Q.4	**(a)**	Discuss the variations of the Apriori algorithm to improve the efficiency.	**07**
	(b)	Describe the various methods which evaluate the accuracy of a classifier or a predictor.	**07**
Q.5	**(a)**	Explain the typical requirements of clustering in data mining.	**07**
	(b)	How the data mining will be used in the retail industry?	**07**
		OR	
Q.5	**(a)**	Write and explain partition clustering algorithm.	**07**
	(b)	What all features should a good data mining system include?	**07**

****