Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

M. E. - SEMESTER – II • EXAMINATION – WINTER • 2013

Subject code: 1720108 **Date: 02-01-2014**

Subject Name: Data Mining & Data warehousing

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		Answer the following.	**14**
	(a)	Define: Knowledge, Information, Data mining	
	(b)	Explain Data warehouse and Data mart.	
	(c)	What are the major challenges in data mining?	
	(d)	Define: Clustering, Outliers.	
Q.2	**(a)**	Explain KDD process with function of each step involved in it.	**07**
	(b)	Why preprocessing is required before any data mining method is applied? Give name of various data preprocessing methods.	**07**
		OR	
	(b)	Define: characterization, discrimination, association, correlation, classification, prediction, evolution analysis.	**07**
Q.3	**(a)**	Explain various OLAP operations performed on Data cube.	**07**
	(b)	What do you mean by data cleaning? Explain various data cleaning methods in brief.	**07**
		OR	
Q.3	**(a)**	Explain following design methods for multidimensional database: Star schema, Snowflake schema, Fact-constellation schema. Also compare them with illustration.	**08**
	(b)	What is application of concept hierarchy? Draw concept hierarchy for location(country, state, city, street) and time(year, quarter, month, week, day).	**06**
Q.4	**(a)**	Compare Apriori and FP-Growth algorithms.	**07**
	(b)	How you differentiate association rule mining from classification? Briefly explain associative classification.	**07**
		OR	
Q.4	**(a)**	Define multilevel association rules. Give suitable example and explain various methods to mine them.	**07**
Q.4	**(b)**	Define: market-basket analysis, transaction dataset, support, confidence, candidate item-set, frequent item-set.	**07**
Q.5	**(a)**	Compare K-mean and K-medoid algorithms with suitable example.	**07**
	(b)	Explain and compare various attribute selection methods for decision tree construction.	**07**
		OR	
Q.5	**(a)**	Explain BIRCH and ROCK algorithms for clustering.	**07**
	(b)	What do you mean by Text mining? Explain various issues involved in it.	**07**

****