Seat No.: _____ *Enrolment No._____*

GUJARAT TECHNOLOGICAL UNIVERSITY

BE SEM-VIII Examination May 2012

Subject code: 183103

Subject Name: Business Intelligence & Data Mining

Date: 08/05/2012 **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)** Explain three-tier Data Warehousing architecture. **07**

(b) Find all frequent itemsets & generate strong association rules in the following database using Apriori Algorithm. Take minimum support count = 2 and min_conf = 70% **07**

TID	Items
T100	I1, I2, I3,I4, I5
T200	I2, I3
T300	I1, I2, I6
T400	I2, I1, I7
T500	I1, I6, I8

Q.2 **(a)** Explain the KDD process in details. **07**

(b) Describe the life cycle of data in detail. **07**

OR

(b) Explain generalization using attribute oriented induction with a complete example. Show necessary steps. **07**

Q.3 **(a)** Suppose that a data warehouse for *Big University* consists of the following four dimensions: *student, course, semester*, and *instructor*, and two measures *count* and *avg_ grade*. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the *avg_grade* measure stores the actual course grade of the student. At higher conceptual levels, *avg_grade* stores the average grade for the given combination. **07**

i. Draw a *snowflake schema* diagram for the data warehouse.

ii. Starting with the base cuboid [*student, course, semester, instructor*], what specific *OLAP operations* (e.g., roll-up from *semester* to *year*) should one perform in order to list the average grade of *CS* courses for each *Big University* student.

iii. If each dimension has five levels (including all), such as "*student* < *major* <*status* < *university* < all", how many cuboids will this cube contain (including the base and apex cuboids)?

(b) Short note: Information gain, logistic regression **07**

OR

Q.3 **(a)** Explain BI/DW architecture in detail. **07**

(b) Describe the following classification methods: **07**

i. CART

ii. Rough Set approach

Q.4	**(a)**	State whether the following statements are true or false. Justify your answers. i. Outliers are always useless. ii. A data warehouse is a database which is maintained separately from an organization's operational databases. iii. OLTP requires updation of data frequently, while it is not required in the data warehouse. iv. In snowflake schema, each dimension is a single highly normalized table.	**07**
	(b)	How can we integrate a Data mining system with a Database or a Data Warehouse?	**07**
		OR	
Q.4	**(a)**	Briefly outline the major steps of decision tree classification.	**07**
	(b)	Difference between OLTP and OLAP systems.	**07**
Q.5	**(a)**	Show one procedure to find frequent itemset without generating the candidate set.	**07**
	(b)	Define the differences between: i. Discrimination and classification. ii. Characterization and clustering. iii. Classification and prediction.	**07**
		OR	
Q.5	**(a)**	Write an essay on "Data mining for Business Intelligence Application".	**07**
	(b)	Describe major issues in data mining.	**07**

****