Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

B.E. - SEMESTER – VIII EXAMINATION – OCTOBER 2012

Subject Code: 183103 **Date: 25/10/2012**

Subject Name: Business Intelligence and Data Mining

Time: 02.30pm - 05.00pm **Total Marks: 70**

Instructions:

1. **Attempt any five questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1 **(a)** Explain Knowledge Discovery in Database process. Why KDD process is popular as Data Mining Process? **07**

(b) Explain how Classification is different from Prediction? Discuss various issues in classification and prediction and write criteria to compare classification and prediction methods. **07**

Q.2 **(a)** On what kind of data, data mining can be applicable? Explain in brief. **07**

(b) Explain the Attribute Oriented Induction in brief. **07**

OR

(b) Explain Concept Description and Data Generalization using specific example. **07**

Q.3 **(a)** Discuss the following terms: **07**

I. Support
II. Confidence
III. Association Rules

(b) A database has set of seven transactions. Each transaction ti is a set of items purchased in a basket in a store by a customer. The set I is the set of all items sold in the store. Let minimum support(min_sup) = 30% and minimum confidence (min_conf) = 80%. Find all frequent item sets using Apriori Algorithm. **07**

TID	List of Items
T1	Beef, Chicken, Milk
T2	Beef, Cheese
T3	Cheese, Boots
T4	Beef, Chicken, Cheese
T5	Beef, Chicken, Clothes, Cheese, Milk
T6	Chicken, Clothes, Milk
T7	Chicken, Milk, Clothes

OR

Q.3 **(a)** Explain FP Growth to generate frequent item sets using specific example. **07**

(b) What are the limitations of Apriori algorithm and how can we increase the efficiency of Apriori Algorithm? **07**

Q.4 **(a)** Define Bayesian Classifier and Bayes Therom. Predict the class label of an unknown sample (given below) using naïve Bayesian Classification for following training data set : **07**

Unknown Sample to classify is :
X = (age = "=30", income = "Medium", Student= "Yes", Credit Rating = "Fair").

Sr. No.	Age	Income	Student	Credit Rating	Class: Buys Computer
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31….40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No
7	31….40	Low	Yes	Excellent	Yes
8	<=30	Medium	No	Fair	No
9	<=30	Low	Yes	Fair	Yes
10	>40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31….40	Medium	No	Excellent	Yes
13	31….40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No

(b) Define tree pruning. Why is tree pruning useful in decision tree induction? How classification rules extracted from decision trees? **07**

OR

Q.4 **(a)** Define Linear, Non-Linear and Multiple Regression. Following table shows the midterm and final exam grades obtained for students in the "Science" subject. **07**

Predict the final exam marks of a student who received 86 marks in the midterm exam using linear regression.

X (Mid Term Marks)	Y (Final Exam Marks)
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

Q.4	**(b)**	Discuss the following terms: I. Neural Network II. Fuzzy Logic III. Rough Set Approach	**07**
Q.5	**(a)**	Define Data Warehouse. Discuss the architecture of Data Warehouse.	**07**
	(b)	Discuss various OLAP operations.	**07**
		OR	
Q.5	**(a)**	Discuss Following: I. Data Mart II. Meta Data III. Business Intelligence	**07**
	(b)	What is the need of Online Analytical Processing (OLAP)? List categories of OLAP tools and explain any one.	**07**

****