Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE – SEMESTER–VIII • Remedial EXAMINATION – WINTER 2013

Subject Code: 183103 Date: 12/09/2013

Subject Name: Business Intelligence and Data Mining

Time: 03:00 pm – 05:30 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)** Describe challenges to data mining regarding data mining methodology and user interaction issues. **07**

(b) Define Data warehouse. Explain terms Data mart and Virtual warehouse. **07**

Q.2 **(a)** Briefly compare snowflake schema, fact constellation and starnet query model. **07**

(b) Discuss various OLAP operations in the multidimensional data model with example. **07**

OR

(b) List out Data cleaning techniques and explain any two. **07**

Q.3 **(a)** What are the differences between the three main types of data warehouse usage: information processing, analytical processing, and data mining? Discuss the motivation behind OLAP mining (OLAM). **07**

(b) Discuss issues to consider during data integration. **07**

OR

Q.3 **(a)** Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. **07**

What is the mean of the data? What is the median?

What is the mode of the data?

What is the midrange of the data?

Can you find roughly the first quartile (Q_1) and the third quartile (Q_3) of the data?

(b) Explain various normalization techniques with example. **07**

Q.4 **(a)** For class characterization, what are the major differences between a data cube based implementation and a relational implementation such as attribute oriented induction? Discuss which method is most efficient and under what conditions this is so. **07**

(b) Consider a database D with 9 transactions as shown in below table **07**

Transaction ID	Item list
T_1	I1,I2,I5
T_2	I2,I4
T_3	I2,I3
T_4	I1,I2,I4
T_5	I1,I3
T_6	I2,I3
T_7	I1,I3
T_8	I1,I2,I3,I5
T_9	I1,I2,I3

Suppose minimum support count is 2 and minimum confidence is 70%. By Applying Apriori algorithm check following rule is accepted or rejected.
Rule1 : I1 ^ I2 => I5
Rule2 : I1 ^ I5 => I2

OR

Q.4 **(a)** Construct FP-tree for following transaction data set. **07**

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

Q.4 **(b)** List and describe the five primitives for specifying a data mining task. **07**

Q.5 **(a)** Brief outline the major steps of decision tree classification. **07**

(b) Discuss various parameters to compare classification method. How to prepare quality data for classification and prediction? **07**

OR

Q.5 **(a)** Explain following classification method: **07**
K-nearest neighbor classifiers
Case-based reasoning

(b) Explain KDD process with neat diagram. **07**

****