Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA - SEMESTER–IV • EXAMINATION – WINTER • 2014

Subject Code: 640005 Date: 06-12-2014

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)** Define following terms: **07**

1. Data Preprocessing
2. Backpropagation
3. Granularity
4. Galaxy Schema
5. Query and Reporting
6. Gini Index
7. Base Cuboid

(b) State whether the following statements are true or false and justify your answer **07**

1. Is Data Mining is another Hype?
2. Data evolution analysis describes and models regularities or trends for objects whose behavior changes over time.
3. Pattern evaluation is an essential process where intelligent and efficient methods are applied in order to extract patterns.
4. A binary variable is symmetric if the outcomes of the states are not equally important, such as the positive and negative outcomes of test.
5. The 0-D Cuboid, which holds the highest level of summarization stated as apex cuboid.
6. The spiral method involves the rapid generation of increasingly functional systems, with short intervals between successive releases.
7. Drill-Down operation performs aggregation on a data cube.

Q.2 **(a)** Suppose that the data for analysis include the attribute age. The age values for the data tuples are (in increasing order): **07**

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.

1. Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.
2. How might you determine outliers in the data?
3. What other methods are there for data smoothing?

(b) In the process of data cleaning, how can we fill up the missing values? Write down its methods. **07**

OR

(b) A database has five transactions. Let min-support=60% & min confidence=80%. **07**

TID	Items-bought
T100	M,O,N,K,E,Y
T200	D,O,N,K,E,Y
T300	M,A,K,E
T400	M,U,C,K,Y
T500	C,O,O,K,I,E

(i) Find all frequent item sets.
(ii) Find all strong association rules.

Q.3 **(a)** Give the syntax of cube and dimension definition for star, snowflake and fact constellation schema by taking any example. **07**

(b) Explain with figure: 3-Tire Data Warehouse Architecture. **07**

OR

Q.3 **(a)** Explain the methods of generating the concept hierarchies for categorical data with suitable example. **07**

(b) Describe Decision Tree Induction algorithm. You can describe it with the help of an example. How are the Rules induced from the Decision Tree? **07**

Q.4 **(a)** List and explain the popular Splitting Rules. **07**

(b) What are the interestingness measures of association rule mining? Explain three interestingness measures giving appropriate examples. **07**

OR

Q.4 **(a)** Explain OLAP Operations in Multi-dimensional Data Model With Examples. **07**

(b) (i) Differentiate Star Schema and Snow-Flake Schema **03**

(ii) List and explain different types of OLAP servers. **04**

Q.5 **(a)** Explain how data mining application is helpful for Telecommunication Industry **07**

(b) Define data reduction? Brief the strategies include in data reduction? **07**

OR

Q.5 **(a)** Describe classifier accuracy measure and explain the term confusion matrix with example **07**

(b) Explain the different types of data repositories on which mining can be performed? **07**

****