

GUJARAT TECHNOLOGICAL UNIVERSITY**MCA - SEMESTER-VI • EXAMINATION – WINTER 2013****Subject Code: 640005****Date: 13-12-2013****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)** Fill in the blanks with four key words appearing in the Data Warehouse definition as given below: **04**
Data warehouse is a _____, _____, _____, and _____ collection of data in support of management's decision making process. Briefly describe these four key words.
- (b)** Briefly describe the following OLAP operations: Roll-up, Drill-down, Slice, Dice, and Pivot **05**
- (c)** What is Classification? What is Clustering? Differentiate these two terms. Data Mining for Financial Analysis includes an application entitled, "Classification and Clustering of Customers for Target Marketing". Briefly explain this application. **05**
- Q.2 (a)** Describe Apriori Algorithm for Rule of Association Mining. You can describe the algorithm using an example. What is Apriori Property? If we have 10 transactions and we go for Item-pairs only, how many pairs will be formed if we assume that all the pairs pass through the minimum support criterion. **07**
- (b)** Suppose that a data warehouse consists of three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Concept hierarchy for time dimension is: day, month, quarter, year; for doctor is doctor, hospital; for patient is patient. **07**
- (i)** Draw a Star Schema for the above data warehouse.
- (ii)** Starting with the base cuboid [month, doctor, patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2010?
- OR**
- (b) (i)** Describe briefly Enterprise Warehouse and Data mart. **07**
- (ii)** Name the three types of OLAP Servers, and briefly describe them
- Q.3 (a)** Filling up Missing Values is one of the aspects of Data Cleaning. What are various methods for filling up Missing Values? Briefly describe them. **05**
- (b)** Describe briefly Data Integration, various related issues and their solution approaches. **05**
- (c)** Pattern Interestingness Measures include (i) Simplicity, and (ii) Certainty. Describe these measures and give reason for their relevance and importance. **04**
- OR**
- Q.3 (a)** Smoothing out Noisy Data is an aspect of Data Cleaning. Describe briefly various methods of Smoothing Noisy Data. **05**
- (b)** Often Data needs to be Transformation so as to make it suitable and convenient for Data Mining. What are various methods of Data Transformation? Describe **05**

them.

- (c) Pattern Interestingness Measures include (i) Utility, and (ii) Novelty. Describe these measures and give reason for their relevance and importance. **04**
- Q.4** (a) What is Attribute-oriented Induction? Briefly describe the meaning and purpose of the 1st step of Attribute-oriented Induction, i.e. Data Focusing. **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**
- OR**
- Q.4** (a) There are two ways to perform Attribute-oriented Induction: (i) Attribute Removal, and (ii) Attribute Generalization. Describe them including two common approaches to control Generalization Process. **07**
- (b) Describe the following methods for estimating and increasing the Accuracy of a Classifier: (i) Holdout Method, (ii) K-fold Cross-validation Method, (iii) Bagging, (iv) Boosting. **07**
- Q.5** (a) Hierarchical Methods of Clustering are usually classified into two approaches: Agglomerative (i.e. Bottom-up), and Divisive (i.e. Top-down). Describe both the approaches of Hierarchical Methods of Clustering. **04**
- (b) K-Means Method is the most well-known and commonly-used Partitioning Method for Clustering of data. Describe it. **06**
- (c) Describe the following applications of Data Mining for Financial Data Analysis: (i) Loan Payment Prediction and Customer Credit Policy Analysis, (ii) Classification & Clustering of Customers for Target Marketing **04**
- OR**
- Q.5** (a) Describe Density-based Methods of Clustering. Do these methods discover clusters of arbitrary shapes? Briefly justify your answer. **04**
- (b) What is an Outlier? What are the potential applications of Outlier Mining? List down various methods of Outlier Detection. **06**
- (c) Describe the following applications of Data Warehousing & Data Mining for the Retail Industry: (i) Multi-dimensional Analysis of Sales, Customers, Products, Time, & Region, (ii) Analysis of Effectiveness of Sales Campaign. **04**
