

Time: 3 Hours

FEBRUARY 2014

Max. Marks: 100

PLEASE WRITE YOUR ROLL NO. AT THE SPACE PROVIDED ON EACH PAGE IMMEDIATELY AFTER RECEIVING THE QUESTION PAPER.

NOTE:

- Question 1 is compulsory and carries 28 marks. Answer any FOUR questions from the rest. Marks are indicated against each question.
- Parts of a question should be answered at the same place.

-
- Q.1**
- What is Data Cleaning? Explain any one Data Cleaning Tool.
 - Distinguish between supervised & unsupervised learning techniques.
 - Explain the need of Data Staging Component.
 - Differentiate between Star Index & Star Join.
 - Describe various guidelines given by E. F. Codd's for OLAP.
 - Explain the importance of Metadata.
 - Enumerate data mining applications for Financial Data Analysis. (7×4)
- Q.2**
- Differentiate between ROLAP & MOLAP on the basis of Data Storage and its various technologies. (9)
 - Design Star Schema with a factless fact table to track patient in hospital by procedure, time & ward type. State all required assumptions. (9)
- Q.3**
- What is the difference between verification & discovery? Explain snapshot and transaction fact tables? Also explain how they are related. (9)
 - Discuss OLAP mining (OLAM). With the help of a diagram discuss architecture of OLAM. (9)
- Q.4**
- Explain Hash-based technique used to improve the efficiency of Apriori-based mining? (6)
 - What is Data Warehouse? Describe 3-tier architecture of Data Warehouse. (8)
 - Describe different criteria based on which classification and prediction methods can be compared and evaluated? (4)

Code: CT75**Subject: DATA WAREHOUSING AND DATA MINING**

- Q.5** a. What is KDD? Discuss the steps involved in KDD processing. (6)
- b. What is data reduction? What are the different strategies for data reduction? (6)
- c. What is outlier mining? Explain the different approach used for statistical based outlier detection. (6)
- Q.6** a. How decision tree assist in Data mining? Explain over- fitting of an induced decision tree and two approaches to avoid over – fitting using suitable example. (9)
- b. How Web can be used as data source for data warehouse. How different type of information can we get from the Web? Explain with example. (9)
- Q.7** Write short notes on the following:
- (i) Drill Down & Roll Up
 - (ii) Genetic Algorithm
 - (iii) Bayesian Classification (6×3)