

**ALCCS – NEW SCHEME**

Time: 3 Hours

**FEBRUARY 2013**

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** a. Define Support factor and Confidence factor with reference to Association rule mining.
- b. Explain briefly ROLAP.
- c. Explain the use of data mining in retail industry citing suitable examples.
- d. Mention few advantages of using Bayesian Networks for data analysis.
- e. Name few techniques to improve the efficiency of Apriori algorithm. Explain briefly one of these.
- f. What are the requirements of cluster analysis? What are the different types of data used for cluster analysis?
- g. Write short notes on multidimensional data model? Define data cube. (7×4)
- Q.2** a. How is a data warehouse different from a database? How are they similar to each other? (9)
- b. Discuss three data warehouse models- the enterprise warehouse, the data mart and the virtual warehouse. (9)
- Q.3** a. Explain briefly star and snowflake schema. Also Point out the major difference between the two. Which is popular in the data warehouse design? (9)
- b. Discuss Data extraction, Data transformation and Data loading with respect to Data warehouse. (9)
- Q.4** a. Discuss typical OLAP operations in brief. (9)
- b. Why most data warehouse systems support index structures? Discuss methods to index OLAP data. (9)



**Code: CT75**

**Subject: DATA WAREHOUSING AND DATA MINING**

- Q.5** 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 OLAM. (8)
- b. Discuss basic algorithm for inducing a Decision tree from training samples. (5)
- c. How is prediction different from classification? List criterion on the basis of which Classification and prediction methods can be compared and evaluated. (5)
- Q.6** a. Discuss various issues in Data mining. (8)
- b. How can efficiency of Apriori-based be improved? Describe briefly any of five variations of the Apriori algorithm. (10)
- Q.7** a. What is Hierarchical method of clustering? Differentiate Agglomerative and Divisive Hierarchical Clustering? (6)
- b. Write short note on any **THREE** of the following:
- (i) Social impacts of data mining
  - (ii) Text mining
  - (iii) Entropy-based discretization
  - (iv) Min-max and z-score normalization
- (3×4)