CMCA/M-20   **10541**

DATA WAREHOUSING AND MINING

Paper–MCA-14-43

Time Allowed : 3 Hours]   [Maximum Marks : 80

Note : Attempt **five** questions in all, selecting at least **one** question from each Unit. Question No. **1** is compulsory. All questions carry equal marks.

Compulsory Question

1. (a) What is data mart and meta data?

    (b) What is Entropy? How entropy is computed?

    (c) Which attribute type, data mining supports?

    (d) Write a brief note on pixel oriented data visualization technique.

    (e) What is meant by table lookup model?

    (f) Write the application areas where density based clustering method are applicable.

    (g) Discuss different types of tree pruning methods.

    (h) Write the efficiency parameters of any three data mining tools.   8×2=16

UNIT–I

2. (a) What are the data computation methods? Differentiate among various categorizations of OLAP tools.

   (b) What are the characteristics of Data warehouse? Outline and explain three tier architecture of Data warehouse. How this architecture is integrated with reporting tools?   8+8=16

3. (a) Write the pre-requisites for data warehouse schema. How fact and dimension tables are identified and designed? Discuss galaxy schema of Data warehouse.

   (b) Explore the different considerations for designing and implementing a data warehouse system.

   8+8=16

UNIT–II

4. (a) What is knowledge discovery in Databases? Explain data mining functionalities with supportive examples.

   (b) "Whether all patterns are interesting"? Comment on the statement and compare between support and confidence.   8+8=16

5. (a) What is Data pre-processing? Explain data integration and data reduction techniques in the pre-processing step of data mining.

(b) What are outliers? Write the different types of outliers. Explain challenges of outlier detection.

8+8=16

UNIT–III

6. Define Clustering? What are the requirements for clustering? How partition clustering is used? Write a detailed note on K-means algorithm along with distance functions. 16

7. (a) What is the nearest neighbouhood? How the performance of nearest neighbourhood is computed? Explore the challenges of memory based reasoning.

(b) What are similarity models? Write the steps for designing similarity models. 8+8=16

UNIT–IV

8. What do you mean by market basket analysis? Discuss the algorithm of apriori for mining association rules in transactional databases with candidate key and without candidate key generation. 16

9. Write detailed note on the following :

(a) Decision Tree Induction.

(b) Bayesian Belief Networks. 8+8=16