Roll No. ..... Total Pages : 04

MCA/M-19 10513

DATA WAREHOUSING AND MINING

MCA-14-43

Time : Three Hours] [Maximum Marks : 80

Note : Attempt *Five* questions in all. Q. No. **1** is compulsory. Attempt *four* more questions selecting *one* question from each Unit.

(Compulsory Question)

1. (a) What are the characteristics of data warehouse ?

(b) Differentiate between MOLAP, ROLAP and HOLAP.

(c) List and explain different types of outliers.

(d) Explain different types of data objects used in data mining process.

(e) List the name of any *three* distance functions which helps in identifying similarities.

(f) Write a short note on lookup table model.

(g) On what basis an attribute is considered as a root in decision tree.

(h) List the name of six important data mining tools depending upon their efficiency. 8×2=16

Unit I

2. (a) What do you know about the time lines of data warehousing development ? Discuss threetier architecture of data warehouse.

   (b) How fact and dimension tables are important for designing data Warehouse Schema ? Draw and explain snow-flake schema. **8+8=16**

3. (a) Draft a 3-D Cube and discuss the various operations which may be applied on the 3-D Cube.

   (b) Discuss the steps for construction and implementation of data warehouse. **8+8=16**

Unit II

4. (a) Define data mining. Discuss data mining functionalities. Draw a sketch of integration from data warehousing to data mining.

   (b) What are the factors which influence the quality of data ? Discuss the strategies of data reduction. **8+8=16**

5. (a) Distinguish between descriptive and predictive data mining models.

   (b) What is data visualization ? How results are easy to interpreter with this methodology ? **8+8=16**

Unit III

6. What do you mean by clustering ? Discuss the typical requirements which are necessary for clustering. Implement hierarchical clustering for the following data points (P) with respect to X-axis and Y-axis.

   P1(2, 2), P2(1, 14), P3(10, 7), P4(1, 11), P5(3, 4), P6(11, 8), P7(4, 3), P8(12, 9) **16**

7. What is nearest neighborhood classification ? Discuss the role of hamming distance for classification of recall pattern. Explore the challenges of memory based reasoning. **16**

Unit IV

8. Design Frequent pattern growth tree (FP-Tree) and identify frequent itemsets using Apriori algorithm from the transactional database as given below using candidate generation with a minimum support threshold of 2(20%).

Transaction	T10	T20	T30	T40	T50	T60	T70	T80	T90	T100
Item ID	11, 12,	12, 14, 15,	12, 13	11, 12, 15	11, 13	12, 13	11, 13	11, 12, 14, 15	11, 12, 13	11, 12, 15

16

9. (a) Define Neural Network. Design a specimen neural network by modelling an OR gate.

(b) "Data mining can be implemented in every sphere of life." Justify this statement with reference to real-life applications, where data mining may be applied. **8+8=16**