Roll No. .....

Total Pages : 3

MCA/M-18 **10513**

DATA WAREHOUSING AND MINING

Paper : MCA-14-43

Time : Three Hours]

[Maximum Marks : 80

Note : Students are required to attempt *five* questions in all. Question No. 1 is compulsory. Students will have to attempt *four* more questions, selecting *one* question from each unit.

Compulsory Question

1. (a) What is information gain? How it is computed?

(b) Differentiate between OLAP and OLTP.

(c) What do you mean by data visualization?

(d) When a pattern is considered to be an interesting?

(e) What do you understand by similarity amongst observation?

(f) How table lookup model is considered as statistical technique?

(g) How tree pre-pruning and post-pruning is implemented in decision tree?

(h) Write a note on support and confidence. (8×2=16)

UNIT–I

2. (a) Describe the historical evolution and paradigm shift of data warehouse systems. Explain the various components of data warehouse architecture.

(b) What,do you mean by data warehouse schemas? How fact and dimension tables are identified and designed? (8×2=16)

3. (a) What is the need of OLAP? Explain various categories of OLAP. How rollup and drill down OLAP operations are implemented?

(b) What are the steps for designing and construction of data warehouse? Explain various considerations. (8×2=16)

UNIT–II

4. (a) Under which criteria a data mining system may be classified. Explain data mining system architecture.

(b) What do you understand by the term data processing? How missing value and noise can be removed? (8×2=16)

5. (a) Explore the different steps to implement directed data mining methodology.

(b) What do you mean by outliers? Discuss different types of outliers with examples. (8×2=16)

UNIT–III

6. What do you mean by clustering? Explain the type of data used in cluster analysis. Discuss K-means 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 neighbourhood classification? Discuss the performance of Nearest Neighbour classifier? How you will differentiate between numeric, non-numeric and mixed attributes.

UNIT–IV

8. What do you understand by association rule mining in transactional databases? Implement the generation of Apriori Algorithm with using candidate key generation for the following dataset with a minimum support threshold of 2(20%).

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

16

9. (a) What do you mean by Neurodes? Explain commonly used Neunet architectures.

(b) Draw a comparative chart of Data mining tool and discuss them in term of accuracy, efficiency and complexity.

(8×2=16)