MCA/ M11
Data Warehousing & Data Mining
Paper : MCA 402

Time: Three Hours **Maximum Marks: 80**

Note: Students will be required to attempt FIVE questions in all. Question No. 1 is compulsory. In addition to compulsory question, students will have to attempt four more questions, selecting one question from each Unit.

1. Write short note on the followings:
   (a) Data Mart
   (b) Loose coupling and Tight coupling
   (c) Join and prune
   (d) Learning by observation
   (e) Slice and Dice Operation
   (f) Continuous and categorical Data
   (g) Item sets
   (h) Binary Variables.

UNIT-I

2. (a) What do you mean by Data Warehouse ? Differentiate between OLTP and OLAP systems.
   (b) What are data cubes? How data cubes can be framed from tables and spreadsheets? Draft a cuboid by taking four dimensions as time, doctor, patient and disease.
2. What do you understand by Data Warehousing schema? Explain the most popular data model for a data warehouse.

UNIT-II

4 a) what do you understand by data mining query language? What are the primitive; under which a data mining query is defined?
   b) What do you mean by data mining? What are the steps of data mining to be followed as the process of knowledge discovery?
5. (a) Why we preprocess the data ? what are the basic methods for filling in the missing value of a database?
   (b) "Attribute-oriented induction generates one or a set of generalized descriptions'. How can these descriptions be visualized?

Unit-III

6. Find the frequent item sets using Apriori algorithm from the transactional database as given below using candidate generation with a minimum support threshold of 2 (20%).

Transaction ID	Item in transaction
T10	3,4
T20	2,3,5
T30	1,2,3,5
T40	2,5
T50	1,2,5
T60	1,3
T70	2,3
T80	1,3
T90	2,3
T100	3,5

7. (a) Discuss the issues regarding classification and prediction.
   (b) Explain classification by Back propagation,

UNIT-IV

8. Differentiate between Data Matrix and Dissimilar Matrix. Explain the areas where data mining clustering may be applied.
9. Under which categories clustering methods are classified? Explain partition based clustering in detail.