MCA – 53

V Semester M.C.A. Examination, Feb./March 2010
DATA MINING AND WAREHOUSING

Time 3 Hrs Max. Marks : 100

Instructions *: 1) Answer **all** questions in Part **A**, **5** out of **8** questions in Part **B** and **4** out of **6** questions in Part **C**.*
*2) Part **A** : Questions from **1** to **8** carry **1** mark **each** and **9** to **14** carry **2** marks **each**.*
*3) Part **B** : **Each** question carries **8** marks.*
*4) Part **C** : **Each** question carries **10** marks*

PART – A

1. What is the different data mining functionalities?
2. Differentiate between operational database systems and data warehouses.
3. Describe five primitives for specifying data mining task.
4. What is classification?
5. Explain concept description characterization.
6. Explain the need of preprocessing.
7. Explain association rule mining with an example.
8. Explain classification by back propagation.
9. Discuss any four OLAP operations in multidimensional data model with examples.
10. With a neat diagram, explain three tire data warehousing architecture.
11. How do you choose a good data mining system?
12. Discuss on trends in data mining.
13. Write a note on analytical characterization.
14. Explain mining single dimensional boolean association rules from transactional databases.

P.T.O.

PART – B

1. Explain the following concepts using an example :
   i) Snow flake schema
   ii) Fact constellation schema iii) Star schema.
2. Explain decision tree classifier.
3. Explain the terms data cleaning and data reduction methods.
4. Differentiate between OLAP and OLTP. Explain the different types of OLAP server Architectures.
5. Explain the four methods of presenting or visualizing the discovered patterns with an example.
6. Describe why concept hierarchies are useful in data mining. Discuss the DMQL syntax for concept hierarchy specification.
7. Explain association rule mining and write apriori algorithm for finding frequent item sets.
8. Explain the K-means and K-mediods classification and partitioning methods.

PART – C

1. Describe the generation of candidate item sets and frequent item sets where the minimum support is 2 for the following transactional data :

T ID	List of item-IDs
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

2. Discuss the Bayesian classification method with suitable example.

3. Give an account on the regression methods used in prediction.

4. a) Describe Linear regression method .

   b) Define Euclidean and Manhattan distance. What are the four requirements of a distance function? If V1= (10,5) and V2 = (8,9), Determine the Euclidean and Manhattan distance values.

5. Explain the terms agglomerative, decisive hierarchical clustering, BIRCH clustering techniques.

6. Give an account on the role of data mining in bio-medical and DNA analysis.

———————