MCA/M-16
DATA WAREHOUSING AND MINING
PAPER- MCA-14-43

Compulsory Question

1. (a) What do you understand by dice and drilldown OLAP operation? State example.

(b) What do you mean by data mart?

(c) Write a note on classification.

(d) Briefly explain data objects and attribute types.

(e) Write a note on table lookup model.

(f) What do you mean eccentric/oddball in nearest neighborhood?

(g) Write a note on join and prune in ARM.

(h) List the name of six data mining tools.

UNIT-I

2. (a) How data generalization may be achieved by attribute oriented induction? Discuss.

(b) What do you mean by materialization in data warehouse implementation? Explain.

3. (a) Discuss three tier architecture of data warehouse. How this architecture integrate to data mining?

(b) What do you mean by fact and dimension tables? How these helps data warehouse schema?

UNIT-II

4. (a) What are the motivations behind data mining? Weather all patterns are interesting in data mining process? Comment.

(b) What do you mean by data visualization? How it is helpful in result interpretation and evaluation?

5. What do you mean by data preprocessing? Why is it important in KDD process?

How data cleaning and reduction is achieved? Discuss.

UNIT-III

6. (a) What do you mean by similarity models? Discuss the steps for designing similarity models?

(b) What are the requirements of clustering? Write the steps to implement K-means algorithm in clustering.

7. Calculate the classification using nearest neighborhood for the following training patterns using Hamming distance function. Assuming hamming distance between attribute is 1 each.

Training Pattern	Attributes			Class
	Course	Game	Dress	
T1	M.Tech.	Cricket	Casuals	Indian
T2	M.Tech.	Foot ball	Formal	Foreign
T3	M.Tech.	Cricket	Formal	Indian
T4	B.Tech.	Hockey	Formal	Foreign
T5	M.Tech.	Hockey	Casuals	Indian
T6	B.Tech.	Cricket	Formal	Foreign
T7	M.Tech.	Hockey	Formal	Indian
T8	B.Tech.	Hockey	Casuals	Foreign

Generate the classification rules for the following recall patterns.

Recall Pattern	Attribute			Class
	Course	Game	Dress	
R1	B.Tech.	Cricket	Casuals	?
R2	B.Tech.	Foot ball	Formal	?
R3	B.Tech.	Foot ball	Formal	?
R4	B.Tech.	Foot ball	Casuals	?

UNIT-IV

8. (a) What do you mean by decision tree? What are the measures for attribute selection? How tree is pruned?

(b) What do you mean by Bayesian Belief Networks? Discuss the potential application areas of Neural Network. How neural networks are modeled?

9. Find the frequent item sets using Apriori algorithm from the transactional database as given below with and without using candidate, generation with a minimum support threshold of 2 (20%).

Transaction ID	Item of Transaction
10	1,2,3
20	1,3,5
30	2,3
40	3,5
50	1,2,5
60	1,3
70	3,4,5
80	2,3,5
90	1,2,3,5
100	2,5