No. of Printed Pages : **2**

Sl. No.

C5-R4 : DATA WAREHOUSING AND DATA MINING

NOTE :

1. **Answer question 1 and any FOUR from questions 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Time : 3 Hours **Total Marks : 100**

1. (a) What is knowledge discovery in database ? How does it is related to data mining ?

(b) Elaborate the following statement. "Dimension table is wide and fact table is deep". Justify your answer with suitable example.

(c) What is meant by Association rule mining ?

(d) Differentiate between ROLAP and MOLAP.

(e) Explain case-based reasoning with suitable example.

(f) What is the difference between view and materialized view ?

(g) What is a decision tree ?

(7x4)

2. (a) What is data preprocessing ? Explain various steps involved in data preprocessing.

(b) Discuss different type of operations that can be performed on data cube.

(c) Explain different component tables of the star schema ? Describe the composition of the primary keys for the dimension and fact tables.

(6+6+6)

3. (a) Differentiate between OLTP and OLAP.

(b) Explain the three-tier data warehouse architecture.

(c) What are three major types of metadata in data-warehouse ? Explain the purpose of each type.

(6+6+6)

4. (a) Explain applications of association rule mining.

(b) Develop the Apriori algorithm for generating frequent item sets.

(c) Consider the following transaction data set :

Tid	1	2	3	4	5	5	7	8	9	10
Items	{a,b}	{b,c,d}	{a,c,d,e}	{a,d,e}	{a,b,c}	{a,b,c,d}	{a}	{a,b,c}	{a,b,d}	{b,c,e}

Construct the FP tree by showing the trees separately after reading each transaction.

(4+6+8)

5. (a) Suppose that the data mining task is to cluster the following eight points, with (x, y) representing location, into 3 clusters.

AI(2, 10); A2(2, 5); A3(8, 4); B1(5, 8); B2(7, 5); B3(6, 4); CI(1, 2); C2(4, 9) :

The distance function is Euclidean distance. Suppose initially, we assign A1, B1 and C1 as the center of each cluster respectively. Use the k-means algorithm to show :

(i) The three cluster centers after the first-round execution

(ii) The final three clusters.

(b) Compute the Euclidean and Manhattan distance between the two objects represented by following tuples (1, 6, 2, 5, 3) and (3, 5, 2, 6, 6).

(12+6)

6. (a) What do you mean by metadata repository ? Why it is required ? What should a metadata repository contain ?

(b) Discuss various applications of data mining.

(c) Explain memory-based reasoning method and its applications.

(6+6+6)

7. (a) For the given confusion matrix below for three classes. Find sensitivity and specificity metrices to estimate predictive accuracy of classification methods.

Predicted Class	True Class		
	1	2	3
1	8	1	1
2	2	9	2
3	0	0	7

(b) How to improve accuracy of classification ? Explain.

(c) Explain Partitioning and Hierarchical methods of cluster analysis.

(10+4+4)

- o O o -