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) Compare OLAP and OLTP.

(b) Define Data Mining. Differentiate between data warehouse and data mart.

(c) Mention eight application areas related with Data Warehousing and Data Mining.

(d) Differentiate between supervised learning and unsupervised learning.

(e) What is the meaning of single linkage and complete linkage in context of hierarchical clustering algorithm ?

(f) Explain confusion matrix for evaluating classifier.

(g) Define following terms with suitable example :

(i) Closed itemsets

(ii) Minimum confidence

(7x4)

2. (a) Explain different varieties of data on which Data Mining may be applied.

(b) Discuss Knowledge Discovery in Databases (KDD) process.

(c) What is data reduction ? What are its importance ? Explain different methods of data reduction.

(6+6+6)

3. (a) Discuss various methods for handling missing values in data cleaning process.

(b) Find mean, median, mode, and range of following data :

6, 4, 5, 2, 5, 7, 4, 8, 7, 4, 1

(c) Define Cuboid ? Explain various OLAP Operations in the context of Data Cube.

(6+4+8)

4. (a) Explain FP - Growth Algorithm for finding Frequent Item - sets.

(b) Write Apriori Property. Generate all frequent itemsets using Apriori algorithm on the following data set with minimum support value as 50%.

TID	Items Purchased
T101	Cheese, Milk, Cookie
T102	Butter, Milk, Bread
T103	Cheese, Butter, Milk, Bread
T104	Butter, Bread

(c) Explain K-means clustering algorithm with example.

(6+6+6)

5. (a) How does DBSCAN works in cluster Analysis ? Explain.
   (b) Explain the types of data in cluster analysis.
   (c) What is Concept Hierarchy ? Explain two types of Concept Hierarchy approaches.

(7+4+7)

6. (a) Mention the names of the schemas used for Multidimensional Databases and further explain any two out of them.
   (b) Differentiate between ROLAP and MOLAP.
   (c) Explain three tier architecture of Data Warehouse using suitable diagram.

(6+4+8)

7. (a) Discuss basic principles of AOI.
   (b) Find out class for p5 using K-nearest neighbor classification algorithm using Euclidean distance measure and considering k=3.

Name	Acid Durability	Strength	Class
p1	7	7	Bad
p2	7	4	Bad
p3	3	4	Good
p4	1	4	Good
p5	3	7	?

   (c) Explain feed forward neural networks in detail.

(4+7+7)

- o O o -