

C5-R4 : DATA WAREHOUSING AND DATA MINING

NOTE :

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

Total Time : 3 Hours

Total Marks : 100

1.
 - (a) Discuss different types of data to which data mining can be applied.
 - (b) What is an attribute ? Explain the different types of attributes with an example.
 - (c) What are data mining task primitives ?
 - (d) How to evaluate the accuracy of a classifier ?
 - (e) Differentiate between supervised and unsupervised learning.
 - (f) Differentiate between ROLAP, MOLAP, and HOLAP.
 - (g) What is data visualization? What are the various ways to visualize data ? (7x4)

2.
 - (a) What is data reduction ? Explain various methods of attribute subset selection using an example.
 - (b) What is data transformation ? Explain the data transformation strategies.
 - (c) What are the value ranges of the following normalization methods ?
 - (i) Min-max normalization
 - (ii) z-score normalization
 - (iii) z-score normalization using mean absolute deviation instead of standard deviation
 - (iv) normalization by decimal scaling (8+6+4)

3.
 - (a) What is a data warehouse ? What are the key features of a data warehouse ? Explain the three-tier architecture of data warehouse.
 - (b) What is a snowflake schema ? Construct a snowflake schema with the following : Sales fact table, with Store, Location, Time, Product, Line, and Family dimension tables. (10+8)

4.
 - (a) What is decision tree classification ? Explain the steps of decision tree classification using a suitable diagram.
 - (b) What are the pros and cons of eager classification and lazy classification ?
 - (c) Explain the K-Means algorithm using a suitable algorithm. (6+6+6)

5. Differentiate between OLAP and OLTP. Explain all the typical operations of OLAP on multidimensional data using a suitable example. Define atleast three dimensions with their aggregate values. (18)

6. A database has five transactions. Let $min_sup=60\%$ and $min_conf=70\%$.
- (a) Find all frequent itemsets using Apriori and FP-growth respectively. Compare the efficiency of the two mining processes.

T_ID	Items Bought
T-1000	M, O, N, K, E, Y
T-1001	D, O, N, K, E, Y
T-1002	M, A, K, E
T-1003	M, U, C, K, Y
T-1004	C, O, O, K, E

- (b) What is case-based reasoning ? Explain with the help of a suitable example. (10+8)
7. (a) What do you mean by data mining for intrusion detection and prevention ?
- (b) What is attribute-oriented induction ? Explain the process in detail.
- (c) What is neural networks? How is back propagation performed ? (5+5+8)

- o O o -