USN ☐☐☐☐☐☐☐☐☐☐   **21IS643**

Sixth Semester B.E./B.Tech. Degree Examination, Dec.2024/Jan.2025
Data Mining and Data Warehousing

Time: 3 hrs.   Max. Marks: 100

Note: *Answer any FIVE full questions, choosing ONE full question from each module.*

Module-1

1   a. Define Datawarehouse. Explain the multi tier architecture of data warehouse.   **(10 Marks)**

    b. Distinguish between OLTP and OLAP systems.   **(10 Marks)**

OR

2   a. Define Data cube. Explain OLAP operations with neat diagram.   **(10 Marks)**

    b. Explain the three schemas for multidimensional data models.   **(10 Marks)**

Module-2

3   a. Define Data mining, with a neat diagram explain the KDD process.   **(10 Marks)**

    b. What are the different strategies that are used in Data preprocessing? Explain Aggregation, sampling and dimensionality reduction with example.   **(10 Marks)**

OR

4   a. Explain the challenges involved in Data mining.   **(10 Marks)**

    b. Explain features subset selection, feature creation and discretization and binarization.

   (10 Marks)

Module-3

5   a. Write the pseudocode for the frequent itemset generation of the Apriori algorithm, explain.   **(10 Marks)**

    b. List and explain the factor that affects the computational complexity of the Apriori algoirhtm.   **(10 Marks)**

OR

6   a. Consider the transaction data set

IID	1	2	3	4	5	6
Items	{a, b}	{b, c, d}	{a, c, d, e}	{a, d, e}	{a, b, c}	{a, b, c, d}}

7	8	9	10
{a}	{a, b, c}	{a, b, d}	{b, c, d}

Construct the FP tree by showing the trees separately after reading each transaction. Also find the frequent itemset ordered by their corresponding suffixes.   **(10 Marks)**

b. Generate frequent item sets and associated rules for the given transaction data set with minsup = 40% and minconf = 70% using Apriori algorithm. **(10 Marks)**

TID	Items bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Bear, Diapers}

Module-4

7 a. With an example, explain the different measures that can be used to determine the best way to split the records. **(10 Marks)**

b. List and explain the different characteristics of decision tree induction. **(10 Marks)**

OR

8 a. Write and explain k-nearest neighbor classification, algorithm. Summarize the characteristics of nearest neighbor classifier. **(10 Marks)**

b. Explain how a rule based classifier works? Also write the sequential covering algorithm to extract rules directly from data. **(10 Marks)**

Module-5

9 a. What is cluster analysis? Describe the different types of clustering techniques and clusters with example. **(10 Marks)**

b. Explain DBSCAN algorithm for clustering with an example. **(10 Marks)**

OR

10 a. Explain basic Agglomerative Hierarchical clustering algorithm. Give its strength and weakness. **(10 Marks)**

b. Write the CLIQUE algorithm. Also explain the strengths and limitations of CLIQUE. **(10 Marks)**

* * * * *