CBCS SCHEME

USN | | | | | | | | | | |

17CS651

Sixth Semester B.E. Degree Examination, Dec.2023/Jan.2024
Data Mining and Data Ware Housing

Time: 3 hrs. Max. Marks: 100

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

Module-1

1 a. What is Data Warehouse? Explain three tier architecture of data warehouse. **(10 Marks)**
b. Explain the schemes of multidimensional data models. **(10 Marks)**

OR

2 a. Explain the categorization of measures. **(04 Marks)**
b. Explain different datawarehouse models. **(06 Marks)**
c. Elaborate on typical OLAP operations on multidimension data with examples. **(10 Marks)**

Module-2

3 a. Briefly explain curse of dimensionality. **(04 Marks)**
b. Identify the different indexing method used for OLAP data with brief explanation. **(08 Marks)**
c. Describe the servers involved in implementation of a warehouse server. **(08 Marks)**

OR

4 a. What is data mining? Explain various data mining tasks, with suitable examples. **(10 Marks)**
b. Define Data preprocessing. Mention the steps involved in it. Explain any two steps in detail. **(10 Marks)**

Module-3

5 a. Describe frequent itemset generation in Apriori algorithm with example. **(10 Marks)**
b. Discuss the different factors affecting computational complexity of the Apriori algorithm. **(10 Marks)**

OR

6 a. Describe the alternative methods for generating frequent item sets. **(10 Marks)**
b. Construct the FP tree for the Table Q6 (b). Show the trees after reading each transaction.

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

Table Q6 (b) **(10 Marks)**

Module-4

7 a. Define Decision tree. Write the algorithm for decision tree induction. **(10 Marks)**

b. Consider the Table Q7 (b) for a binary classification problem.

Instance	a_1	a_2	a_3	Target class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	–
4	F	F	4.0	+
5	F	T	7.0	–
6	F	T	3.0	–
7	F	F	8.0	–
8	T	F	7.0	+
9	F	T	5.0	–

Table Q7 (b)

(i) What is the entropy of this collection of training examples with respect to the positive class?

(ii) What are the information gains of a_1 and a_2 relative to the training examples?

(iii) For a_3, which is a continuous attribute, compute the information gain for every possible split. **(10 Marks)**

OR

8 a. What is rule-based classifier? Explain sequential covering algorithm in rule-based classifier. **(10 Marks)**

b. Write an algorithm for K-nearest neighbor (KNN) classification. Summarize the characteristics of nearest neighbor classifiers. **(10 Marks)**

Module-5

9 a. What is Cluster analysis? Discuss the different types of clusters with examples. **(10 Marks)**

b. Perform a hierarchical clustering of five samples using the single linkage algorithm shown in Table 9(b) and two features X and Y. Draw the Dendogram.

	1	2	3	4	5
X	4	8	15	24	24
Y	4	4	8	4	12

Table Q9 (b) **(10 Marks)**

OR

10 a. Discuss DBSCAN algorithm for clustering. **(10 Marks)**

b. Write DENCLUE algorithm for Kernel density estimation. Briefly explain strengths and limitations of the same. **(10 Marks)**

* * * * *