CBCS SCHEME

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

17CS651

Sixth Semester B.E. Degree Examination, July/August 2022
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. What is Data Warehouse? What are its features? **(06 Marks)**

b. Explain with diagram, a three tire data warehouse architecture. **(08 Marks)**

c. Compare OLTP and OLAP systems. **(06 Marks)**

OR

2 a. What is Metadata? Explain its features. **(06 Marks)**

b. Explain the following terms with examples:
Star Schema, Snowflake schema and Constellation schema. **(09 Marks)**

c. List the major functions involved in ETL process. **(05 Marks)**

Module-2

3 a. With respect to indexing explain bitmap index and join index. **(08 Marks)**

b. What is data mining? Explain KDD process in data mining. **(08 Marks)**

c. For the following vectors X and Y, calculate the similarity, where
X = {0, 1, 0, 1}
Y = {1, 0, 1, 0} **(04 Marks)**

OR

4 a. What is data preprocessing? Explain various data pre-processing tasks. **(07 Marks)**

b. Describe ROLAP, MOLAP and HOLAP. **(09 Marks)**

c. For the following vectors X and Y, calculate the Jaccard coefficient, where
X = {0, 1, 0, 1}
Y = {1, 0, 1, 0} **(04 Marks)**

Module-3

5 a. What is Apriori principle? Write and explain apriori algorithm for frequent itemset generation. **(08 Marks)**

b. Briefly explain the candidate generation procedure using $F_{k-1} \times F_{k-1}$ method. **(06 Marks)**

c. Generate frequent itemset for the given data with support = 50%.

TID	1	2	3	4
Items	{1, 3, 4}	{2, 3, 5}	{1, 2, 3, 5}	{2, 5}

(06 Marks)

OR

6 a. What is FP growth algorithm? In what way it is used to find frequent itemsets. **(08 Marks)**

b. Explain the uses of Hash tree in support counting. **(06 Marks)**

c. Consider the following transaction dataset. Describe the construction of FP-tree in FP growth algorithm, assuming min-support as 2.

Tid	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}

(06 Marks)

Module-4

7 a. How decision trees are used for classification? Explain decision tree induction algorithm for classification. **(10 Marks)**

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

OR

8 a. Explain k-nearest neighbor classification algorithm with an example. **(07 Marks)**

b. Explain sequential covering algorithm in rule-based classifier. **(07 Marks)**

c. Write a note on Bayesian classifier. **(06 Marks)**

Module-5

9 a. What is cluster analysis? Explain different types of clustering. **(10 Marks)**

b. What are the basic approaches used for generating agglomerative hierarchical clustering? **(10 Marks)**

OR

10 a. Explain DBSCAN algorithm with example. **(08 Marks)**

b. Explain k-means clustering algorithm in brief. **(06 Marks)**

c. Define SSE. What are the strategies used for reducing SSE in k-means clustering algorithm. **(06 Marks)**

* * * * *