CBCS SCHEME

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

18CS641

Sixth Semester B.E. Degree Examination, June/July 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. What is Data Warehouse? Explain Three –tier Data Warehousing Architecture. **(06 Marks)**
b. Compare OLTP and OLAP systems. **(04 Marks)**
c. Explain the Schemas and Multidimensional Data Models. **(10 Marks)**

OR

2 a. Explain the OLAP operations in Multidimensional Data Models. **(10 Marks)**
b. What is Data cube measure? Explain the categorization of measures. **(10 Marks)**

Module-2

3 a. Explain different Indexing methods on OLAP data. **(08 Marks)**
b. Explain the curse of dimensionality and Data Cube Materialization. **(06 Marks)**
c. Explain the implementation of OLAP server architecture. **(06 Marks)**

OR

4 a. What is Data Mining and explain various Data Mining tasks with example. **(10 Marks)**
b. For the following vectors X and Y, calculate the cosine similarity, where
X = { 3, 2, 0, 5, 0, 0, 0, 2, 0, 0 }
Y = { 1, 0, 0, 0, 0, 0, 0, 1,0,2] **(04 Marks)**
c. Define Data Preprocessing. Mention the stages involved in it. Explain any two steps in detail. **(06 Marks)**

Module-3

5 a. Write the Apriori algorithm for frequent item set generation. **(06 Marks)**
b. Explain the Rule- generation in Apriori algorithm. **(06 Marks)**
c. Illustrate the advantages of using closed frequent Item sets with an example, show the relationship among frequent, Maximal frequent and closed frequent item sets. **(08 Marks)**

OR

6 a. Construct an FP- Tree for the following dataset:

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 }

(10 Marks)

b. Explain various methods for generating frequent itemsets. **(10 Marks)**

Module-4

7 a. Write an algorithm for Decision Tree Induction and briefly explain the characteristics of Decision Tree Induction. **(10 Marks)**
b. Consider the following dataset for a Binary classification problems:

Interval	a1	a2	a3	Target Clss
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	-

i) What is the Entropy of this collection of training example with respect to the positive class?
ii) What are the information gins of a1 and a2 relative to these training examples.
iii) For a3, 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. Explain Naïve Bayes classifier, Write the characteristics of Naïve Bayes Classifier. **(10 Marks)**

Module-5

9 a. What is cluster analysis? Describe the different types of clusters. **(10 Marks)**
b. State and explain K-means algorithm. **(10 Marks)**

OR

10 a. Explain DB scan algorithm with example. **(10 Marks)**
b. Explain Agglomerative flierarchical clustering Algorithm with different proximity between clusters. **(10 Marks)**

* * * * *