



CBCS SCHEME

18CS641

Sixth Semester B.E./B.Tech. Degree Examination, Dec.2025/Jan.2026 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. Explain the various OLAP operations with a neat diagram. (08 Marks)
- b. Compare OLTP and OLAP systems. (08 Marks)
- c. Describe the multitiered architecture of a Data Warehouse System. (04 Marks)

OR

- 2 a. Define the following terms: (08 Marks)
 - (i) Data cube
 - (ii) Metadata
 - (iii) Dimension
 - (iv) Facts
- b. How can measures be categorized? (08 Marks)
- c. What are the steps to be followed when designing a Data Warehouse System? (04 Marks)

Module-2

- 3 a. Explain Bitmap and Join Indexing with appropriate examples. (08 Marks)
- b. Summarize the various challenges that motivated the development of data mining. (08 Marks)
- c. Explain the various ways of achieving data cube materialization. (04 Marks)

OR

- 4 a. Explain the following terms : (08 Marks)
 - (i) Wrapper Approach
 - (ii) Filter Approach
 - (iii) Outlier
 - (iv) Precision
- b. Describe ROLAP, MOLAP and HOLAP. (08 Marks)
- c. Compute SMC and Jacard similarity coefficient for (04 Marks)

$$X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$$

18CS641

Module-3

- 5 a. Explain the following candidate generation procedures :

- (i) Brute Force
- (ii) $F_{K-1} * F_1$
- (iii) $F_{K-1} * F_{K-1}$

(10 Marks)

- b. Write the steps for frequent itemset generation of Apriori algorithm. (05 Marks)

- c. List the frequent itemsets using Apriori for support count = 30%.

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

(05 Marks)

OR

- 6 a. Draw the FP Tree for the given data :

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. What are the alternative methods of generating frequent itemsets? (10 Marks)

Module-4

- 7 a. Explain the characteristics of Naïve Bayes classifiers. (10 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and/or equations written eg. 42+8 = 50, will be treated as malpractice.

CMRIT LIBRARY
BANGALORE - 560 037

- b. Explain Hunts algorithm to develop a decision tree. Consider the following training set and derive the decision tree.

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125 K	No
2	No	Married	100 K	No
3	No	Single	70 K	No
4	Yes	Married	120 K	No
5	No	Divorced	95 K	Yes
6	No	Married	60 K	No
7	Yes	Divorced	220 K	No
8	No	Single	85 K	Yes
9	No	Married	75 K	No
10	No	Single	90 K	Yes

(10 Marks)

OR

- 8 a. Explain K-nearest neighbor classification algorithm. Apply KNN and predict the type of food to which Tomato (Sweet = 6, Crunch = 4) belongs.

Ingredient	Sweet	Crunch	Food type
Grape	8	5	Fruit
Green Bean	3	7	Vegetable
Nuts	3	6	Protein
Orange	7	3	Fruit

(10 Marks)

- b. Explain Bayes Theorem. Consider a football game between two teams : Team 0 and Team 1. Team 0 wins 65% of the time and Team 1 wins the remaining matches. Among the games won by Team 0, only 30% of them come from playing on Team 1's field. 75% of victories for Team 1 are obtained while playing at home. If Team 1 is to host the next match between the two teams, which team will most likely emerge as the winner? (10 Marks)

Module-5CMRIT LIBRARY
BANGALORE - 560 037

- 9 a. Explain DBSCAN Algorithm. (10 Marks)
- b. Explain : (i) K-Means Clustering
(ii) Graph Based Clustering (10 Marks)

OR

- 10 a. Explain Balanced Iterative Reducing and Clustering using hierarchies. (10 Marks)
- b. Write a note on CURE. (10 Marks)
