



USN

--	--	--	--	--	--	--	--	--	--

18CS641

**Sixth Semester B.E. Degree Examination, June/July 2024**  
**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 architecture of data warehouse with a neat diagram. (10 Marks)
- b. Explain the various differences between OLTP and OLAP systems. (10 Marks)

**OR**

- 2 a. Explain with suitable examples the various OLAP operations in a multi-dimensional data model. (10 Marks)
- b. Explain the concept of star, snowflake and fact constellation schemas for multi-dimensional databases with examples. (10 Marks)

**Module-2**

- 3 a. What is data mining? With a neat diagram, explain KDD process in data mining. Also explain the various data mining tasks with suitable examples. (12 Marks)
- b. For the following vectors X and Y, calculate
  - (i) Cosine similarity
  - (ii) Euclidean distance
  - (iii) Jaccard coefficient
  - (iv) Correlation coefficient
$$X = (0, 1, 0, 1); Y = (1, 0, 1, 0)$$

(08 Marks)

**OR**

- 4 a. With respect to indexing, explain bit map index and join index operations. (10 Marks)
- b. Briefly explain any five data preprocessing approaches. (10 Marks)

**Module-3**

- 5 a. What is frequent itemset generation? Explain frequent itemset generation using Apriori principle with an example. (10 Marks)
- b. Describe alternative methods for generating frequent itemsets. (10 Marks)

**OR**

- 6 a. Briefly explain the factors affecting the computational complexity of apriori algorithm. (04 Marks)
- b. What is association analysis? Explain association rule, support and confidence with an example. (06 Marks)
- c. Explain objective measures of interestingness for evaluating association patterns. (10 Marks)

**Module-4**

- 7 a. Define classification. With a neat diagram explain the general approach to solve classification problem. (08 Marks)
- b. Illustrate Hunt's algorithm to develop a decision tree. Using hunt's algorithm derive decision tree for the following data :

	Binary	Categorical	Continuous	Class
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

(12 Marks)

**OR**

- 8 a. What is a rule based classifier? Explain the following :  
 (i) Sequential covering algorithm.  
 (ii) Rule ordering schemes. (10 Marks)
- b. Write an algorithm for K-nearest neighbour classification. List the characteristics of nearest neighbor classifiers. (10 Marks)

**Module-5**

- 9 a. What is Cluster analysis? Explain the different types of clustering techniques with examples. (10 Marks)
- b. Explain K-means clustering algorithm. What are its limitations? (10 Marks)

**OR**

- 10 a. Explain Agglomerative hierarchical clustering algorithm with different proximity between clusters. (10 Marks)
- b. Explain DBSCAN algorithm and estimate time and space complexity. How the parameters are selected? (10 Marks)

\*\*\*\*\*