

# CBCS SCHEME

USN

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

18CS641

Sixth Semester B.E. Degree Examination, Dec.2023/Jan.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. Differentiate between operational database systems and data warehouse. (10 Marks)  
b. Explain the data warehouse models. (10 Marks)

OR

- 2 a. Explain typical OLAP operations with examples. (10 Marks)  
b. What are the schemas of multidimensional data models? Explain. (10 Marks)

### Module-2

- 3 a. List different indexing methods for OLAP data and explain with examples. (10 Marks)  
b. Differentiate between ROLAP, MOLAP and HOLAP servers. (10 Marks)

OR

- 4 a. What is data mining? Explain KDD process with the help of neat diagram. (10 Marks)  
b. Define similarity and dissimilarity between the objects. Find SMC and Jaccard's coefficient of two binary vectors given below.  
 $X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$   
 $Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$ . (06 Marks)  
c. For the following two data objects, calculate cosine similarity.  
 $x = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0)$   
 $y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2)$ . (04 Marks)

### Module-3

- 5 a. Define the Apriori principle. Explain frequent item set generation in the Apriori Algorithm. (10 Marks)  
b. What is association analysis? Define association rule, support and confidence by giving an example for each. (10 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 objective measures of interestingness for evaluation of quality of association patterns. (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.

**Module-4**

- 7 a. How does decision tree algorithm work. Explain with example. (10 Marks)  
b. Examine the methods for comparing classifiers. (10 Marks)

OR

- 8 a. Describe Nearest Neighbor classifier. List its characteristics. (10 Marks)  
b. Explain Bayesian classifier with an example. (10 Marks)

**Module-5**

- 9 a. What is cluster analysis? Discuss the different types of clusters with examples. (10 Marks)  
b. Explain Agglomerative hierarchical clustering. How do you define proximity between clusters? (10 Marks)

OR

- 10 a. Discuss DBSCAN algorithm with an example. (10 Marks)  
b. Explain the following :  
i) Density based clustering  
ii) Graph based clustering. (10 Marks)

\*\*\*\*\*