

# CBCS SCHEME

17CS651



## Sixth Semester B.E. Degree Examination, Feb./Mar. 2022 Data Mining and Data Warehousing

Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

### Module-1

- 1 a. Define a Data Warehouse. With a neat diagram, describe the 3 – tier architecture of a Data Warehouse. (10 Marks)
- b. Compare OLTP and OLAP. (06 Marks)
- c. What are the key features of Data Warehouse? (04 Marks)

OR

- 2 a. With examples, explain the three schemas for multidimensional data models. (09 Marks)
- b. What is Data Cube measure? Explain the categories of the measure. (07 Marks)
- c. Explain Roll up and drill down operations with examples. (04 Marks)

### Module-2

- 3 a. Explain the different OLAP server architecture. (06 Marks)
- b. What is Data Mining? What are the two major categories of Data Mining Task? (06 Marks)
- c. Explain the four Core Data Mining Tasks. (08 Marks)

OR

- 4 a. Explain different types of Attributes with example. (04 Marks)
- b. What is Dimensionality Reduction? Explain how it can be achieved using feature subset solution. (08 Marks)
- c. Define Similarity and dissimilarity between objects. Compute SMC and Jaccard's coefficient for the following vector :  
 $X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$      $Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$ . (08 Marks)

### Module-3

- 5 a. What is Association Rule Mining? Explain the terms support and confidence with examples. (06 Marks)
- b. List the factors that affect the computational complexity of Apriori algorithm. (06 Marks)
- c. Explain Support counting using Hash structure. (08 Marks)

OR

- 6 a. Construct an FP tree for the following data set and describe the steps of construction.

TID	Items
1	{a, b}
2	{b, c, d}
3	{a, c, d, e}
4	{a, d, e}
5	{a, b, c}

(10 Marks)

- b. Explain Candidate generation using  $F_1 \times F_{k-1}$  and  $F_{k-1} \times F_{k-1}$  merging strategies. (10 Marks)

**Module-4**

- 7 a. Describe the general approach to solving a classification problem and also discuss how the model is evaluated. (10 Marks)  
b. Explain the characteristics of decision tree induction algorithms and also comment on the design issues in decision tree induction. (10 Marks)

OR

- 8 a. Describe the Nearest Neighbor classifier. (06 Marks)  
b. What are the characteristics of Nearest Neighbor classifier? (06 Marks)  
c. Explain the Bayesian classifier. (08 Marks)

**Module-5**

- 9 a. What is Clustering? Discuss the different types of Clustering. (08 Marks)  
b. Write and explain K means algorithm. (08 Marks)  
c. Discuss the issues with K – means Clustering. (04 Marks)

OR

- 10 a. Write a note on the following : (06 Marks)  
i) Prototype based cluster      ii) Graph based cluster. (08 Marks)  
b. Explain DBSCAN clustering in detail. (06 Marks)  
c. What are the important issues in Cluster Validation? (06 Marks)

CMRIT LIBRARY  
BANGALORE - 560 037

\*\*\*\*\*