

III B. Tech I Semester Regular Examinations, Dec/Jan – 2022-23**DATA WAREHOUSING & DATA MINING**

(CSE-IOT&CS Including Block Chain Technology)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions **ONE** Question from **Each unit**

All Questions Carry Equal Marks

UNIT-I

1. a) What is a Data warehouse? List out and explain the common sectors where Data warehouse is used. [7M]
 b) Explain the characteristics and benefits of OLAP for Businesses. [7M]

(OR)

2. a) Describe the major components of Data Warehouse. [7M]
 b) Explain the characteristics of Star schema in Data Warehouse modeling. [7M]

UNIT-II

3. a) Define Data Mining. Discuss its applications and advantages. [7M]
 b) Present various methods for Data Normalization with suitable examples. [7M]

(OR)

4. a) Why is Data preprocessing an inevitable phase in KDD process? Explain the necessary steps and tasks of data preprocessing. [7M]
 b) What is the importance of Similarity and Dissimilarity measures in data analytics? Briefly discuss various Similarity and Dissimilarity measures. [7M]

UNIT-III

5. How is the FP growth algorithm solving the pattern mining problem? Explain the step-by-step procedure. [14M]

(OR)

6. a) Discuss the two types of Classification using frequent patterns. [7M]
 b) Explain the Constraint based frequent pattern mining. [7M]

UNIT-IV

7. a) What are the measures used for selecting attributes in decision tree induction algorithm? Explain some popular attribute selection measures. [7M]

- b) What is SVM? How SVM is used for classification? [7M]

(OR)

8. a) Explain the operational concepts of k-Nearest Neighbor classification. And specify how to choose the best value of 'k'? [7M]

- b) Discuss various techniques to improve Accuracy of Classification models. [7M]

UNIT-V

9. a) Discuss various internal and external measures used for cluster evaluation. [7M]
b) Demonstrate k-means clustering technique and also discuss its strengths and weaknesses. [7M]
- (OR)
10. a) Describe the four types of linkages used in Hierarchical clustering. [7M]
b) Explain the working principle of CURE algorithm for clustering high dimensional data. [7M]