

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY GURAJADA VIZIANAGARAM**  
**III B. Tech I Semester Regular/Supplementary Examinations, April/May -2025**  
**DATA WAREHOUSING AND DATA MINING**  
**(CSE)**

Time: 3 hours

Max. Marks: 70

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

All Questions Carry Equal Marks

\*\*\*\*\*

| <b>UNIT-I</b>   |    |  |      |
|-----------------|----|--|------|
| 1.              | a) | Define OLAP and explain the different OLAP operations used in data cube computation with suitable examples.                    | [7M] |
|                 | b) | Discuss the differences between star schema, snowflake schema, and fact constellation schema in data warehouse modeling.       | [7M] |
|                 |    | (OR)   |      |
| 2.              | a) | Describe the functionalities and characteristics of data mining systems.   | [7M] |
|                 | b) | Explain the steps in designing and implementing a data warehouse.  | [7M] |
| <b>UNIT-II</b>  |    |  |      |
| 3.              | a) | Illustrate the steps of data cleaning with examples.   | [7M] |
|                 | b) | What is data discretization? Compare and contrast binning, histogram analysis, and clustering methods used for discretization. | [7M] |
|                 |    | (OR)   |      |
| 4.              | a) | Explain the role and techniques of data integration in the context of heterogeneous data sources.                              | [7M] |
|                 | b) | Discuss various methods for data reduction and how they help improve performance in data mining.                               | [7M] |
| <b>UNIT-III</b> |    |  |      |
| 5.              | a) | Describe the process of building a decision tree using the Gini index. Include a step-by-step example.                         | [7M] |
|                 | b) | What is overfitting in decision trees? How does pre-pruning and post-pruning help mitigate it?                                 | [7M] |
|                 |    | (OR)   |      |
| 6.              | a) | Explain the concept of scalability in classification. How is scalability addressed in decision tree algorithms?                | [7M] |
|                 | b) | Compare classification and prediction. How is accuracy measured in classification models?                                      | [7M] |
| <b>UNIT-IV</b>  |    |  |      |
| 7.              | a) | Explain confidence-based pruning in association rule generation. Why is it necessary?  | [7M] |
|                 | b) | Demonstrate how the FP-Growth algorithm works with a simple transactional dataset.   | [7M] |
|                 |    | (OR)   |      |
| 8.              | a) | What are closed and maximal frequent itemsets? Discuss their advantages in association analysis                                | [7M] |
|                 | b) | Compare and contrast Apriori and FP-Growth algorithms in terms of computational complexity.                                    | [7M] |
| <b>UNIT-V</b>   |    |  |      |
| 9.              | a) | Define clustering. Explain the characteristics of hierarchical clustering and its types.                                       | [7M] |
|                 | b) | Illustrate how K-means clustering works with a numerical example.  | [7M] |
|                 |    | (OR)   |      |
| 10.             | a) | Compare density-based and partitioning-based clustering techniques. In which scenarios is DBSCAN preferred over K-means?       | [7M] |
|                 | b) | Discuss the applications of clustering in bioinformatics and market basket analysis.   | [7M] |

\*\*\*\*\*