R22 Regulation *Subject code:4E6FB*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech VI Semester Supplementary Examinations, November 2025

DATA WAREHOUSING AND DATA MINING

(IT)

Maximum Marks: 60 Date: 19.11.2025 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 10 marks. Answer all questions in Part A.
3. Part B consists of 5 Units. Answer any one full question from each unit.
4. Each question carries 10 marks and may have a, b, c, d as sub questions.

Part-A

All the following questions carry equal marks (10X1M=10 Marks)		Marks	CO	Bloom Tx
1.a)	How is data ware house different from a database?	1M	CO1	II
b)	What is data transformation?	1M	CO1	I
c)	List the steps involved in the process of KDD.	1M	CO2	I
d)	What is data cleaning?	1M	CO2	I
e)	Quote an example for quantitative association rule.	1M	CO3	I
f)	List the advantages of FP growth algorithm.	1M	CO3	II
g)	How do you evaluate the accuracy of a classifier?	1M	CO4	I
h)	What is the main idea of naive Bayesian classification?	1M	CO4	I
i)	Define outliers.	1M	CO5	I
j)	Give the objectives of clustering.	1M	CO5	I

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	Bloom Tx
2	Diagrammatically illustrate and discuss the 3-tier data warehouse architecture and its various components.	10M	CO1	IV
	OR			
3	Discuss the characteristics of OLAP cube and different OLAP operations with an example.	10M	CO1	III
4	Why do we need to preprocess the data? Discuss the different forms of preprocessing with examples.	10M	CO2	III
	OR			
5	a) Present an account of the various data mining issues and functionalities in detail. b) Classify the KDD process and explain.	5M 5M	CO2	III
6	Assume 5 transactions and explain the two step approach to generate frequent item sets and to mine association rules using Apriori algorithm.	10M	CO3	IV
	OR			

7	a) Write the step-by-step approach to generate frequent patterns using FP growth algorithm. b) Generate association rules from the frequent item set $\{I_1, I_2, I_5\}$ by assuming 0.8 confidence.	6M 4M	CO3	V
8	Describe the Decision tree induction algorithm for classification. Discuss the usage of information gain in this.	10M	CO4	III
	OR			
9	Discuss about k-nearest neighbor classification algorithm and its characteristics with an example.	10M	CO4	III
10	a) Classify various Clustering methods. b) How does partitioning around medoids algorithm achieve the goals of clustering?	4M 6M	CO5	III
	OR			
11	Explain in detail about the types of Hierarchical Clustering.	10M	CO5	III