R22 Regulation *Subject code: 4E5EC*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech V Semester Regular/Supplementary Examinations, November 2025

DATA WAREHOUSING AND DATA MINING

(CSE)

Maximum Marks: 60 Date:10.11.2025 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)	Define the Fact table.	1M	1	1
b)	What is ROLAP server?	1M	1	2
c)	Write about binning technique.	1M	2	2
d)	Define min-max normalization.	1M	2	2
e)	Define frequent itemset.	1M	3	1
f)	What is association rule?	1M	3	2
g)	Define classification problem.	1M	4	1
h)	Write Bayes theorem.	1M	4	1
i)	Define Cluster.	1M	5	1
j)	Define outlier.	1M	5	1

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	Bloom Tx
2	a) Explain the differences between operational database systems and Data warehouse. b) Explain with a neat diagram, multi-tiered architecture of data warehouse.	5M 5M	1	2
	OR			
3	a) Suppose that a data warehouse for Big University consists of the four dimensions student, course, semester, and instructor, and two measures count and avg grade. At the lowest conceptual level, the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination. Draw a snowflake schema diagram for the data warehouse. b) Explain OLAP operations with examples.	5M 5M	1	3
4	a) Explain the steps involved in the KDD process with a neat diagram. b) Discuss about data cleaning techniques with examples.	5M 5M	2	

	OR			
5	a) Discuss on various numerosity reduction techniques. b) Explain measuring data similarity and dissimilarity between different types of attributes with examples.	5M 5M	2	2
6	Explain Apriori algorithm. Consider the following dataset TID List of item IDs T100 I1, I2, I5 T200 I2, I4 T300 I2, I3 T400 I1, I2, I4 T500 I1, I3 T600 I2, I3 T700 I1, I3 T800 I1, I2, I3, I5 T900 I1, I2, I3 Apply Apriori algorithm to generate candidate itemsets and frequent itemsets, where the minimum support count is 2.	10M	3	3
	OR			
7	a) Construct FP Tree for the above transaction dataset. b) Discuss Maximal, Closed Frequent Item Set with examples.	5M 5M	3	3
8	a) Explain decision tree with an example. b) Discuss the measures for selecting best split in decision tree.	5M 5M	4	3
	OR			
9	a) Explain how Bayes theorem is used for classification. b) Discuss KNN classification algorithm and discuss characteristics of nearest neighbor classifiers.	5M 5M	4	2
10	a) Explain different types of clusters. b) Discuss K Means algorithm.	5M 5M	5	2
	OR			
11	a) Discuss Agglomerative hierarchical clustering algorithm. b) Explain issues in hierarchical clustering.	5M 5M	5	2

2-7 B