R22 Regulation *Subject code:4E5EC*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech V Semester Supplementary Examinations, April 2026

DATA WAREHOUSING AND DATA MINING

(CSE)

Maximum Marks: 60 Date:22.04.2026 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	BTL
1.a)	What is data mart?	1	1	1
b)	Explain MOLAP server.	1	1	1
c)	What is dimensionality reduction?	1	2	2
d)	Explain the importance of histogram.	1	2	2
e)	What is confidence measure?	1	3	1
f)	Discuss the advantages of FP-Growth algorithm.	1	3	2
g)	Explain how to split binary attributes.	1	4	2
h)	Discuss Bayes theorem.	1	4	1
i)	What are different types of cluster algorithms?	1	5	2
j)	What are strengths of hierarchical clustering?	1	5	2

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
2	a) Distinguish OLTP and OLAP systems. b) Sketch Data warehouse Architecture and explain its Components.	5M 5M	1	2
	OR			
3	a) Distinguish the differences between Star and Snowflake schema architectures with example. b) Explain OLAP operations with suitable examples.	5M 5M	1	2
4	a) Briefly explain the steps in the KDD process with a neat diagram. b) Discuss methods to handle missing and noisy data during data cleaning.	5M 5M	2	2
	OR			
5	a) Given the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25,25,30,33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52,70. (i) Use min-max normalization to transform the value 35 for age onto the range [0.0,1.0].	6M	2	2

	(ii) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years. (iii) Use normalization by decimal scaling to transform the value 35 for age. b) Explain methods for attribute subset selection.	4M		
6	Explain FP-Growth algorithm. Consider the transaction data-set: TransID Items T1 {a,b} T2 {b,c,d} T3 {a,c,d,e} T4 {a,d,e} T5 {a,b,c} T6 {a,b,c,d} T7 {a} T8 {a,b,c} T9 {a,b,d} T10 {b,c,e} Construct the FP tree by showing the trees separately after reading each transaction.	10M	3	3
	OR			
7	a) Explain Prune and Join operations used in Apriori algorithm. b) Explain Maximal, Closed Frequent Item Set with examples.	5M 5M	3	2
8	a) Explain methods for expressing attribute test conditions for different attribute types. b) Discuss Decision Tree Induction algorithm.	5M 5M	4	2
	OR			
9	a) Construct a Bayesian belief network for detecting heart disease and heartburn in patients. b) Explain nearest-neighbor classifiers and write KNN classification algorithm.	5M 5M	4	3
10	a) Explain PAM algorithm. b) Discuss different types of clusters.	6M 4M	5	2
	OR			
11	a) Discuss Agglomerative hierarchical clustering algorithm with example. b) Explain key issues in hierarchical clustering.	5M 5M	5	3