



R18 Regulation

Subject code:2P6EC

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Supplementary Examinations, December 2024

DATA WAREHOUSING AND DATA MINING (CSE)

Maximum Marks: 70

Date:07.01.2025

Duration: 3 hours

- Note:
- 1.This question paper contains two parts A and B.
 2. Part A is compulsory which carries 20 marks. Answer all questions in Part A.
 3. Part B consists of 5 Units. Answer any one full question from each unit which carries 10M.
 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 (10X2M=20) Marks)		CO	Bloom Tx
1	Define Data Warehousing.	1	L1
2	List out the operations of OLAP	1	L1
3	What is Data Mining?	2	L1
4	Define Correlation analysis.	2	L1
5	Define APRIORI Principle	3	L1
6	What is Maximal Frequent Item Set?	3	L1
7	What is Classification?	4	L1
8	Define Information Gain	4	L1
9	Define Clustering.	5	L1
10	Discuss on Agglomerative and divisive clustering techniques	5	L1

Part-B

Answer All the following questions.		(5X10M=50Marks)	CO	Bloom Tx												
11	What is the Significance of OLAP in Data warehouse.? Describe OLAP operations with neat diagram/example. [10M]		1	L2												
	OR															
12	Explain with suitable diagrams Star, Snow-Flake and Fact constellation Schema. [10M]		1	L2												
13	Explain the various steps involved in KDD. [10M]		2	L2												
	OR															
14	Write short notes on A) Dimensionality Reduction (5M) B) Feature subset selection (5M)		2	L2												
15	A database has five transactions. Let $min_sup = 60\%$ and $min_conf = 80\%$		3	L2												
	<table><tr><td>TID</td><td>Items_bought</td></tr><tr><td>T100</td><td>{M, O, N, K, E, Y}</td></tr><tr><td>T200</td><td>{D, O, N, K, E, Y}</td></tr><tr><td>T300</td><td>{M, A, K, E}</td></tr><tr><td>T400</td><td>{M, U, C, K, Y}</td></tr><tr><td>T500</td><td>{C, O, O, K, I, E}</td></tr></table>	TID	Items_bought	T100	{M, O, N, K, E, Y}	T200	{D, O, N, K, E, Y}	T300	{M, A, K, E}	T400	{M, U, C, K, Y}	T500	{C, O, O, K, I, E}			
TID	Items_bought															
T100	{M, O, N, K, E, Y}															
T200	{D, O, N, K, E, Y}															
T300	{M, A, K, E}															
T400	{M, U, C, K, Y}															
T500	{C, O, O, K, I, E}															

	Find all Frequent item sets using Apriori algorithm. List all the strong association rules. [10M]		
	OR		
16	A) Explain how can you improve the performance of Apriori algorithm. [5M] B) What are the advantages of FP-Growth algorithm? [5M]	3	L2
17	Explain Decision tree induction algorithm for classification. Discuss the usage of information gain in this. [10M]	4	L2
	OR		
18	Why is naive Bayesian classification called “naive”? Briefly outline the major ideas of naive Bayesian classification. Explain Naive-Bayes classification. [10M]	4	L2
19	What are the different Clustering methods? Explain in detail. [10M]	5	L2
	OR		
20	A) Write K-means Clustering algorithm. [5M] B) Write the Key issue in hierarchical clustering algorithm. [5M]	5	L2