



R18 Regulation Subject code:2P6EC  
**TKR COLLEGE OF ENGINEERING AND TECHNOLOGY**  
(Autonomous, Accredited by NAAC with 'A+' Grade)

**B.Tech VI Semester Supplementary Examinations, July 2024**

**DATA WAREHOUSING AND DATA MINING**  
(CSE)

Maximum Marks: 70

Date:30.07.2024 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	Highlight any one advantage of Data Preprocessing in data mining		2	L1
4	Define Binaryzation.		2	L1
5	Define Association rule mining two step processes.		3	L1
6	Give a note on Closed Frequent Item Set.		3	L1
7	How prediction is different from classification?		4	L1
8	How effective are Bayesian classifiers?		4	L2
9	Define Clustering.		5	L2
10	Define PAM.		5	L2

Part-B

Answer All the following questions.		(5X10M=50Marks)		
11	What is the Significance of OLAP in Data warehouse.? Describe OLAP operations with neat diagram/example. [10M]		1	L1
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	L3												
	<table border="1"> <thead> <tr> <th>TID</th> <th>Items bought</th> </tr> </thead> <tbody> <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> </tbody> </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 [5] b. What are the advantages of FP-Growth algorithm? [5]	3	L3												
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 is the goal of clustering? How does partitioning around medoids algorithm achieve this goal? [10M]	5	L4												
	OR														
20	What are the different clustering methods? Explain in detail. [10M]	5	L3												