



R20 Regulation

Subject code:3P6EB

**TKR COLLEGE OF ENGINEERING AND TECHNOLOGY**

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

**B.Tech VI Semester Supplementary Examinations, May 2025**

**DATA WAREHOUSING AND DATA MINING**

(CSE)

Maximum Marks: 70

Date: 18.06.2025

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.
  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)		Marks	CO	BTL
1	What are the characteristics of star schema?	2M	1	L1
2	List any 4 differences between OLTP and OLAP.	2M	1	L1
3	What are the different forms of data preprocessing?	2M	2	L1
4	Define KDD.	2M	2	L1
5	Define support and confidence in Association rule mining with formulae.	2M	3	L1
6	Mention the importance of association rule mining.	2M	3	L1
7	How prediction is different from classification?	2M	4	L1
8	Write a short note on confusion matrix?	2M	4	L1
9	Define clustering.	2M	5	L1
10	What are the requirements for cluster analysis?	2M	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL																								
11	Draw and explain 3-tier architecture of a data warehouse.	10M	1	L2																								
OR																												
12	List and explain OLAP operations with an example.	10M	1	L2																								
13	Perform Decimal scaling normalization on sal attribute in below dataset.	10M	2	L2																								
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>ID</th> <th>GENDER</th> <th>AGE</th> <th>SAL</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>F</td> <td>27</td> <td>19000</td> </tr> <tr> <td>2</td> <td>M</td> <td>51</td> <td>64</td> </tr> <tr> <td>3</td> <td>M</td> <td>52</td> <td>10000</td> </tr> <tr> <td>4</td> <td>F</td> <td>33</td> <td>55000</td> </tr> <tr> <td>5</td> <td>M</td> <td>45</td> <td>45000</td> </tr> </tbody> </table>	ID	GENDER	AGE	SAL	1	F	27	19000	2	M	51	64	3	M	52	10000	4	F	33	55000	5	M	45	45000			
ID	GENDER	AGE	SAL																									
1	F	27	19000																									
2	M	51	64																									
3	M	52	10000																									
4	F	33	55000																									
5	M	45	45000																									
OR																												
14	Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000 a) min-max normalization by setting min=0 and max=1 b) z-score normalization	5M 5M	2	L2																								

15	<p>Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=60% confidence=80%</p> <table border="1"> <thead> <tr> <th>Tid</th> <th>items</th> </tr> </thead> <tbody> <tr> <td>T1</td> <td>{M,O,N,K,E,Y}</td> </tr> <tr> <td>T2</td> <td>{D,O,N,K,E,Y}</td> </tr> <tr> <td>T3</td> <td>{M,A,K,E}</td> </tr> <tr> <td>T4</td> <td>{M,U,C,K,Y}</td> </tr> <tr> <td>T5</td> <td>{C,O,O,K,I,E}</td> </tr> </tbody> </table>	Tid	items	T1	{M,O,N,K,E,Y}	T2	{D,O,N,K,E,Y}	T3	{M,A,K,E}	T4	{M,U,C,K,Y}	T5	{C,O,O,K,I,E}	10M	3	L2
Tid	items															
T1	{M,O,N,K,E,Y}															
T2	{D,O,N,K,E,Y}															
T3	{M,A,K,E}															
T4	{M,U,C,K,Y}															
T5	{C,O,O,K,I,E}															
OR																
16	<p>Discuss compact representation of frequent itemset (closed and maximal) with min-sup=3</p> <table border="1"> <thead> <tr> <th>Tid</th> <th>items</th> </tr> </thead> <tbody> <tr> <td>T1</td> <td>{A,B,C,D}</td> </tr> <tr> <td>T2</td> <td>{A,B,C,D}</td> </tr> <tr> <td>T3</td> <td>{A,B,C}</td> </tr> <tr> <td>T4</td> <td>{B,C,D}</td> </tr> <tr> <td>T5</td> <td>{C,D}</td> </tr> </tbody> </table>	Tid	items	T1	{A,B,C,D}	T2	{A,B,C,D}	T3	{A,B,C}	T4	{B,C,D}	T5	{C,D}	10M	3	L2
Tid	items															
T1	{A,B,C,D}															
T2	{A,B,C,D}															
T3	{A,B,C}															
T4	{B,C,D}															
T5	{C,D}															
17	Describe the general approaches to solve a classification problem.	10M	4	L2												
OR																
18	a) Explain how evaluation of classifiers will be done using confusion matrix.	5M	4	L2												
	b) List the advantages and disadvantage of using decision trees.	5M														
19	a) Write about evaluation of clustering algorithms?	5M	5	L2												
	b) Write the key issues in hierarchical clustering algorithm?	5M														
OR																
20	Discuss key issues involved in hierarchical and K-Means clustering.	10M	5	L2												