



R20 Regulation

Subject code: 3P6GB

# TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Supplementary Examinations, November 2025

## DATA WAREHOUSING AND DATA MINING

(CSE(AI&ML))

Maximum Marks: 70

Date: 21.11.2025

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	Differentiate OLAP, ROLAP and HOLAP.	2M	1	L1
2	Give an example for snow-flake schema.	2M	1	L1
3	List the various forms of data preprocessing.	2M	2	L1
4	What is the need of confidence measure in association rule mining?	2M	2	L1
5	Define Association rule mining two step processes.	2M	3	L1
6	Write short note on support and confidence measures.	2M	3	L1
7	What are the classification techniques?	2M	4	L1
8	Write decision tree algorithm.	2M	4	L1
9	Give a note on k - means algorithm.	2M	5	L1
10	What are the weaknesses of hierarchical clustering?	2M	5	L1

### Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
11	Explain the 3-tier data warehouse architecture and its various components.	10M	1	L2
OR				
12	Design a star schema, snow-flake schema and fact-constellation schema for the data warehouse that consists of the following four dimensions: (Time, Item, Branch and Location). Include the appropriate measures required for the schema.	10M	1	L2
13	Briefly explain the data pre-processing techniques in data mining.	10M	2	L2
OR				
14	How will you handle missing value in a dataset before mining process?	10M	2	L2
15	A database has five transactions. Let min_sup=80% and min sup 3, min_confidence=100% TID LIST_OF_ITEMS T1 {M,O,N,K,E,Y} T2 {D,ON,K,E,Y} T3 {M,A,K,E} T4 {M,U,C,K,Y} T5 {C,O,O,K,I,E}	5M 5M	3	L2

	a) Generate frequent itemsets using apriori algorithm. b) List all the strong association rules.																														
	OR																														
16	Explain FP-Growth Algorithm and construct FP- Growth Tree Support threshold=50%, Confidence= 60% Table: <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Transaction</th> <th>List of items</th> </tr> </thead> <tbody> <tr> <td>T1</td> <td>I1,I2,I3</td> </tr> <tr> <td>T2</td> <td>I2,I3,I4</td> </tr> <tr> <td>T3</td> <td>I4,I5</td> </tr> <tr> <td>T4</td> <td>I1,I2,I4</td> </tr> <tr> <td>T5</td> <td>I1,I2,I3,I5</td> </tr> <tr> <td>T6</td> <td>I1,I2,I3,I4</td> </tr> </tbody> </table> Support threshold=50% and min_sup=3.	Transaction	List of items	T1	I1,I2,I3	T2	I2,I3,I4	T3	I4,I5	T4	I1,I2,I4	T5	I1,I2,I3,I5	T6	I1,I2,I3,I4	10M	3	L2													
Transaction	List of items																														
T1	I1,I2,I3																														
T2	I2,I3,I4																														
T3	I4,I5																														
T4	I1,I2,I4																														
T5	I1,I2,I3,I5																														
T6	I1,I2,I3,I4																														
17	Describe the data classification process with a neat diagram. How does the NaiveBayesian classification works? Explain.	10M	4	L2																											
	OR																														
18	a) State Bayes theorem How can it be applied for data classification? b) With example explain Bayesian belief network.	5M 5M	4	L2																											
19	Suppose that the data mining task is to cluster the following eight students into three clusters, the distance function is Manhattan. Assign record 1,2,3 as the centroid of each cluster respectively. Use the k-means algorithm to show the final three clusters.	10M	5	L2																											
	<table border="1" style="margin-left: 20px;"> <thead> <tr> <th>RecordID</th> <th>Height(cms)</th> <th>Weight(kgs)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>145</td> <td>35</td> </tr> <tr> <td>2</td> <td>165</td> <td>55</td> </tr> <tr> <td>3</td> <td>170</td> <td>90</td> </tr> <tr> <td>4</td> <td>135</td> <td>60</td> </tr> <tr> <td>5</td> <td>140</td> <td>50</td> </tr> <tr> <td>6</td> <td>160</td> <td>75</td> </tr> <tr> <td>7</td> <td>150</td> <td>40</td> </tr> <tr> <td>8</td> <td>155</td> <td>65</td> </tr> </tbody> </table>	RecordID	Height(cms)	Weight(kgs)	1	145	35	2	165	55	3	170	90	4	135	60	5	140	50	6	160	75	7	150	40	8	155	65			
RecordID	Height(cms)	Weight(kgs)																													
1	145	35																													
2	165	55																													
3	170	90																													
4	135	60																													
5	140	50																													
6	160	75																													
7	150	40																													
8	155	65																													
	OR																														
20	Appraise the importance of outlier detection and its application. Explain any one approach for outlier detection.	10M	5	L2																											