R18 Regulation *Subject code:2P6EC*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Regular/Supplementary Examinations, June 2022

DATA WAREHOUSING AND DATA MINING

(CSE)

Maximum Marks: 70 Date:22.06.2022 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)

1 Define metadata with an example
2 Compare the features of OLTP and OLAP
3 Examine about knowledge discovery in data mining.
4 Justify the practical importance of dimensionality reduction
5 What are the drawbacks of apriori algorithm?
6 Discuss about the applications of association rule mining
7 Differentiate classification and prediction
8 How bayesian belief network works?
9 Show the ways to measure the cluster quality
10 Write the functions of PAM clustering algorithm.

Part-B

Answer All the following questions. (5X10M=50Marks)

11 Explain Roll up, Drill down, slice, dice and pivot operation w.r.t OLAP. [10]

OR

12 Demonstrate data warehouse architecture. Explain all its components. [10]

13 How to do data mining tasks like preprocessing, handling missing data, cleaning data and feature selection for student's database. [10]

OR

14 A. Define similarity measures. [5]
B. Show the working of Euclidean and Manhattan distance measures. [5]

15 A. Discuss about basic concepts of frequent itemset mining. [5]
B. Explain Aprori Algorithm. [5]

OR

16 A. Write about FP Growth Algorithm. [5]
B. List down the advantages of FP Growth algorithm. [5]

17 A. Explain classification algorithm with naïve-bayes classifier. [5]

B. Demonstrate decision tree algorithm with weather prediction example. [5]

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

OR

18 Discuss about K-nearest neighbor algorithm with an example. How to calculate nearest neighbor? [10]

19 What are the advantages of hierarchical clustering techniques? Explain any one agglomerative technique in detail. [10]

OR

20 A. Compare K-means and PAM algorithms. [5]

B. Write notes on Outlier detection with an example. [5]