R18 Regulation *Subject code:* 2P5FC

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech V Semester Regular Examinations, February 2021

DATA WAREHOUSING & DATA MINING

(Information Technology)

Maximum Marks: 70 Date: 22.02.2021 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)**

1 Define data warehouse and data mart.
2 What is the use of metadata in data warehouse?
3 What is the major task done in data pre-processing?
4 How will you measure similarity and dissimilarity?
5 Define support and confidence.
6 What is Pivot in data cube.
7 What are the classification techniques?
8 What are the methods for expressing attribute test conditions?
9 Define outliers. How will you determine outliers in the data?
10 List out the difference between characterization and clustering.

Part-B

Answer All the following questions. **(10M X 5=50Marks)**

11 Explain the 3-tier data warehouse architecture and its various components. (10M)

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)

13 Briefly explain the data pre-processing techniques in data mining. (10M)

OR

14 How will you handle missing value in a dataset before mining process? (10M)

15 Describe the process of generating association rules from the identified frequent itemsets. (10M)

OR

16 Explain the steps in the decision tree classification process with an example. (10M)

17 Why naïve Bayesian classification is called naïve? Briefly outline the major ideas of naïve Bayesian classification. (10M)

OR

18 Explain the process of multi-layer feed-forward neural network classification using back propagation learning. (10M)

19 Explain k-means algorithm with an example. Describe the pros and cons of k-means in comparison with the k-medoids algorithm. (10M)

OR

20 A. Explain the working principle of divisive clustering with example. (5M)
B. What are the disadvantages of agglomerative hierarchical clustering algorithm? (5M)