R18 Regulation *Subject code:* **2P5FC**

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech V Semester Supplementary Examinations, July 2022

DATA WAREHOUSING & DATA MINING

(Information Technology)

Maximum Marks: 70 Date:05.07.2022 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.
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 List out the operations of OLAP.
2 List the Data warehouse Characteristics.
3 What is predictive mining?
4 Mention any three measures of Similarity
5 Write the purpose of Apriori algorithm.
6 How to compute confidence measure for an association rule?
7 What is rule classification?
8 Define decision tree.
9 What are the time and space complexities of K-means clustering algorithm?
10 Compare agglomerative and divisive methods.

Part-B

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

11 A. Make a comparisons between the MOLAP and HOLAP. [5]
B. Discuss the star and snowflake schema in detail with suitable example. [5]

OR

12 A. With a neat sketch, Explain three tier architecture of data ware housing. [5]
B. Explain various data warehouse models. [5]

13 A. Describe the steps involved in data mining in the process of KDD. [5]
B. Explain different data mining Functionalities. [5]

OR

14 A. Discuss briefly about data cleaning techniques. [5]
B. Illustrate the Data Transformation by Normalization. [5]

15 A. Explain the procedure to mining closed frequent data item sets. [5]
B. Explain Apriori algorithm with example. [5]

OR

16 A. Write FP- growth algorithm. [5]
B. Explain how association rules are generated from frequent item sets. [5]

17 A. Discuss K- Nearest neighbor classification-Algorithm and Characteristics. [5]
B. How to evaluate the classifier accuracy? [5]

OR

18 Explain decision tree induction algorithm for classifying data tuples and discuss suitable example. [10]

19 A. Give a brief note on PAM Algorithm. [5]
B. What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? [5]

OR

20 What are outliers? Discuss the methods adopted for outlier detection. [10]