R17 Regulation *Subject code:* 1P7FA

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech IV Year I Semester Regular Examinations, February 2021

DATA WAREHOUSING & DATA MINING

(Information Technology)

Maximum Marks: 70 Date:20.02.2021 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 warehousing.
2 List the applications of OLAP.
3 What are the ways to fill missing data?
4 Write the equation for cosine similarity, Euclidian distance and mention its necessity in data mining.
5 How do you calculate support and confidence in association rule mining?
6 Examine the advantages of Frequent Pattern Growth Algorithm.
7 Compare prediction and classification with example
8 Justify the importance Naïve bayes algorithm in classification
9 Identify the situations where K-means clustering algorithm is not suitable
10 Define outlier? Identify outliers in the temperature dataset {21˚C, 28 ˚C,-4 ˚C, 32 ˚C, 5 ˚C,34 ˚C }

Part-B

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

11 A. Sketch the data warehouse architecture. (5M)
B. Explain Star schema and snowflake schema with example. (5M)

OR

12 A. Compare OLAP and OLTP. (5M)
B. Demonstrate ROLAP and MOLAP with example. (5M)

13 A. What are the various phases in data mining? (5M)
B. When to use dimensionality reduction? (5M)

OR

14 A. Discuss the different types of transformations in data mining. (5M)
B. Analyze about various similarity measures used for data analysis. (5M)

15 A. Differentiate Maximal Frequent item set and closed frequent item set. (5M)
B. Show the working principle of FP growth algorithm. (5M)

OR

16 A. Explain the uses of association rule mining with example. (5M)
B. Explain in detail about apriori algorithm. (5M)

17 A. How to evaluate classification algorithm? (5M)
B. Demonstrate the decision tree algorithm with example. (5M)

OR

18 A. List down various classification algorithms. (5M)
B. Classify COVID-19 dataset into +ve(1) or –ve(0) with attributes{ fever, body pain, breathing trouble, drycough, loss of taste or smell} using Suitable classification algorithm. (5M)

19 A. Explain about DBSCAN Algorithm. (5M)
B. How to fix centroid points in Partitioning clustering algorithms? (5M)

OR

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