R20 Regulation *Subject code: 3P4HA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech IV Semester Regular/Supplementary Examinations, September 2023

Data Warehousing & Data Mining

(CSE (DATA SCIENCE))

Maximum Marks: 70 Date:15.09.2023 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 Write the limitations of Data Warehouse.
2 List out the OLAP operations.
3 Define data cleaning.
4 What is difference between Dimensionality reduction and Feature subset selection?
5 What is an association? Write a short note about association rule mining.
6 What are the drawbacks of FP growth algorithm?
7 State classification.
8 What is Splitting Criterion?
9 Explain the different types of data used in clustering.
10 State the hierarchical methods.

Part-B

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

11 Write the differences between operational databases and data warehousing. (10M)

OR

12 Discuss briefly about multi-dimensional data models. (10M)
13 Explain major challenges in data mining. (10M)

OR

14 A. Describe about Data discretization. (5M)
B. Write about Dimensionality reduction methods. (5M)
15 Explain the partition algorithm with an example. (10M)

OR

16 A. Can we overcome the draw backs of APRIORI algorithm? Discuss. (5M)
B. Discuss about basic concepts of frequent itemset mining. (5M)
17 Explain about the classification and discuss with an example. (10M)

OR

18 Discuss in detail about the decision tree induction algorithm. (10M)

19 Explain in brief about PAM algorithm with suitable examples. (10M)

OR

20 Discuss about the outliers. Explain the weakness and strengths in hierarchical clustering methods. (10M)