R18 Regulation Subject code: 2P6EC

# TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

## B.Tech VI Semester Supplementary Examinations, January 2023

## DATA WAREHOUSING AND DATA MINING

*(Computer Science and Engineering)*

***Maximum Marks: 70*** 21-01-2023 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.
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 any two differences between OLTP and OLAP systems.
2 What are the characteristics of Star Schema?
3 Highlight any one advantage of Data Preprocessing in data mining.
4 Define Binaryzation.
5 Define support and Confidence measure.
6 Give a note on closed frequent Itemset
7 What is classification?
8 Define information Gain.
9 Compare Agglomerative and Divisive Clustering techniques.
10 List the various types of clustering methods.

Part-B

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

11 a) Give examples for defining star, snowflake and fact constellation schemas. (5M)

b) Discuss about a three-tier data warehouse architecture. (5M)

OR

12 What is the significance of OLAP in data warehouse? Describe OLAP operations with necessary diagram/example. (10M)

13 Explain different Data Mining tasks for knowledge discovery. (10M)

OR

14 Explain the various data preprocessing techniques.? How data reduction helps in data processing. (10M)

15 a) State and explain Apriori Algorithm. (5 M)

b) Find the frequent itemset from the given data using Apriori algoritham. Take minimum support = 2. (5M)

| TID | List of Items IDs |
|---|---|
| T1 | I1, I3, I4 |
| T2 | I2, I3, I5 |
| T3 | I1, I2, I3, I5 |
| T4 | I2, I5 |

OR

16 Find frequent itemset using FP growth method. (10M)
Take minimum support = 2.

| T ID | ITEMS |
|---|---|
| T100 | I1, I2, I5 |
| T200 | I2, I4 |
| T300 | I2, I3 |
| T400 | I1, I2, I4 |
| T500 | I1, I3 |
| T600 | I2, I3 |
| T700 | I1, I3 |
| T800 | I1, I2, I3, I5 |
| T900 | I1, I2, I3 |

17 Explain Decision tree induction algorithm for classification. Discuss the usage of information gain in this. (10M)

OR

18 Why is naive Bayesian classification called "naive"? Briefly outline the major ideas of naive Bayesian classification. Explain Naive-Bayes classification. (10M)

19 What are the different Clustering methods? Explain in detail. (10M)

OR

20 a) Write K-means Clustering algorithm. (5M)
b) Write the Key issue in hierarchical clustering algorithm. (5M)