



R18 Regulation

Subject code: 2P6EC

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Supplementary Examinations, June/July 2023

DATA WAREHOUSING AND DATA MINING

(Computer Science and Engineering)

Maximum Marks: 70

Date:03.07.2023 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 Define Data Warehousing
- 2 List out the operations of OLAP
- 3 What is Data Mining?
- 4 Define Binaryzation.
- 5 Define APRIORI Principle.
- 6 What are Maximal Frequent Item Set?
- 7 Define predictive modelling.
- 8 What are Bayesian Belief Networks?
- 9 Write on 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 What is the Significance of OLAP in Data warehouse.? Describe OLAP operations with neat diagram/example. [10]

OR

- 12 Discuss the architecture of dataware house system and its components. [10]
- 13 Explain the various steps involved in KDD. [10]

OR

- 14 Explain the various data preprocessing techniques.? How data reduction helps in data processing. [10]
- 15 A database has five transactions. Let $min\ sup = 60\%$ and $min\ conf = 80\%$ [10]

TID	Items bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all Frequent item sets using Apriori algorithm. List all the strong association rules.

OR

- 16 Explain how can you improve the performance of Apriori algorithm [10]

- 17 Explain decision tree induction algorithm for classifying data tuples and discuss suitable example. [10]
OR
- 18 Why is naive Bayesian classification called “naive”? Briefly outline the major ideas of naive Bayesian classification. Explain Naive-Bayes classification. [10]
- 19 a) Give a brief note on PAM Algorithm. (5M)
b) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? (5M)
OR
- 20 What are the different clustering methods? Explain in detail. [10]