R20 Regulation *Subject code: 3P6EB*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Regular/Supplementary Examinations, July 2024

Data Warehousing and Data Mining

(Computer Science and Engineering)

Maximum Marks: 70 Date:22.07.2024 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)		CO	Bloom Tx
1	Write the key components of a data warehouse architecture?	1	L1
2	What are OLAP operations?	1	L1
3	Define Data Mining?	2	L1
4	list at least three primary objectives of data mining process.	2	L2
5	Define the term "frequent item set" in the context of data mining.	3	L1
6	What is Apriori algorithm?	3	L1
7	list three common classification technique in data mining?	4	L1
8	What is Naive Bayes classifier?	4	L1
9	List the methods of clustering.	5	L1
10	What is k-means algorithm?	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)			
11	A) Explain the differences between operational database systems and data warehouses with examples. [5] B) List the benefits and limitations of ROLAP, MOLAP, and HOLAP architectures. [5]	1	L2
	OR		
12	A) Write the advantages and disadvantages of using star schema vs. snowflake schema in data warehouse design. [5] B) Explain a multi-dimensional data model with an example. [5]	1	L2
13	A) Explain the concept of Data Preprocessing and provide two examples of preprocessing techniques. [5] B) Explain the differences between KDD and Data mining. [5]	2	L2
	OR		
14	A) Analyze the differences between Dimensionality Reduction and Feature Subset Selection techniques, providing examples of when each might be used. [5] B) Explain the effectiveness of various measures of Similarity and Dissimilarity in Data Mining. [5]	2	L4

15	A) Explain the concept of support and confidence in association rule mining and describe their significance. [5] B) Write the steps involved in Apriori algorithm with an example. [5]	3	L3
	OR		
16	A) Examine the partition algorithm and compare its approach to that of the Apriori algorithm. [5] B) Assess the benefits and limitations of using compact representations of frequent item sets, such as maximal and closed frequent item sets. [5]	3	L4
17	A) Explain the general approaches to solving a classification problem and illustrate with an example. [5] B) Analyze the methods for expressing attribute test conditions in decision tree construction and discuss their significance. [5]	4	L4
	OR		
18	Design a decision tree for a binary classification problem, including the steps for attribute selection and tree construction with an example. [10]	4	L2
19	A) Apply the k-means algorithm to a small dataset to cluster the data points and explain your process. [5] B) Use the agglomerative hierarchical clustering method to cluster a sample dataset and describe the steps involved. [5]	5	L3
	OR		
20	A) Examine the key issues in hierarchical clustering and compare the strengths and weaknesses of agglomerative and divisive methods. [5] B) Evaluate the effectiveness of the PAM (Partitioning Around Medoids) algorithm compared to the k-means algorithm and justify their use in different scenarios. [5]	5	L4