R20 Regulation *Subject code: 3P6GB*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

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

DATA WAREHOUSING AND DATA MINING

(CSE(AI&ML))

Maximum Marks: 70 Date:24.06.2023 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 State the characteristics of OLAP cube.
2 Which schema is snowflake?
3 What are the limitations of data mining?
4 State the need for data cleaning
5 What are closed and maximal item sets?
6 What is the Apriori property?
7 How do you choose best split while considering a decision tree?
8 What are the pros of decision trees?
9 Why are outliers important in data mining?
10 Differentiate clustering and classification

Part-B

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

11 A. Diagrammatically illustrate and discuss the data warehousing architecture. 5M
B. In Data warehouse technology, a multiple dimensional view can be implemented by a ROLAP, MOLAP or HOLAP. For each technique explain how each of the following functions may be implemented
i. The generation of a data warehouse (including aggregation)
ii. Roll-up
iii. Drill-down
iv. Incremental updating
Which implementation techniques do you prefer and why? 5M

OR

12 Explain KDD process and data pre-processing techniques in detail. 10M

13 A. What is data mining and knowledge discovery? Illustrate data mining tasks with examples. 5M
B. List out some specific challenges that motivated the development of data mining. 5M

OR

14 A. What do you mean by data reduction techniques? Discuss attribute subset selection method with the help of suitable example. 5M
B. How do you measure similarity & dissimilarity of data objects? 5M

15 A. Giving concrete example, explain a method that performs frequent item set mining by using the prior knowledge of frequent item set properties 5M
B. Summarize in detail about various kinds of association rules 5M

OR

16 A. Find all frequent item sets for the following data using FP-growth algorithm with minimum support=60% and minimum confidence =80% 5M

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}

B. What is a frequent item set? How to find frequent item sets for a transactional database? Explain any one approach with illustrations. 5M

17 A. Build a Decision tree for the given example and derive the rules 10M

TID	Refund	Marital status	Taxable income	Cheat
1	Yes	Single	135K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	Yes	Single	90K	Yes

OR

18 A. Why naive Bayesian classification is called naïve? Briefly discuss the ideas of naïve Bayesian classification 5M

B. What is the purpose of "Attribute selection measures" in classification by decision tree induction? How we can use the "Tree pruning" in classification? 5M

19 Apply k-means algorithm and find centroids. Take A & B as a initial centroids
A(1,2), B(9,10), C(8,6), D(5,5), E(3,3), F(6,1),G(2,2) H(2,8) 10M

OR

20 A. With an example discuss dendogram representation for hierarchical clustering of data objects. 5M

B. Outline basic K- means algorithm. Why this algorithm is extended? Explain any one extension. 5M