R20 Regulation *Subject code:3P4HA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech IV Semester Supplementary Examinations, July 2024

Data Warehousing and Data Mining

(CSE(DS))

Maximum Marks: 70 Date:20.07.2024 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	List and define the characteristics of data warehouse.	1	L1
2	Define metadata.	1	L1
3	Define concept Hierarchy generation.	2	L1
4	Define the process of KDD.	2	L1
5	Define support and confidence in Association rule mining with formulae.	3	L1
6	Mention the importance of association rule mining.	3	L1
7	Define two steps in classification.	4	L1
8	Define root node, internal node and leaf node.	4	L1
9	Define clustering.	5	L1
10	Differentiate clustering and classification.	5	L1
	Part-B		
Answer All the following questions. (5X10M=50Marks)			
11	Draw and explain 3-tier architecture of a data warehouse (10M)	1	L2
	OR		
12	List and explain OLAP operations with an example(10M)	1	L3
13	a) Given the following measurements for the variable age (5M) 18,22,25,42,28,43,33,35,56,28 Standardize the variable by using Z-score normalization. b) Explain data reduction methods. (5M)	2	L2
	OR		
14	a) Briefly discuss the data smoothing methods with an example. (5M) b) Explain how to handle missing values with an example. (5M)	2	L2
15	Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=60% confidence=80% (10M)	3	L2

	<table><tr><td>Tid</td><td>items</td></tr><tr><td>T1</td><td>{M,O,N,K,E,Y}</td></tr><tr><td>T2</td><td>{D,O,N,K,E,Y}</td></tr><tr><td>T3</td><td>{M,A,K,E}</td></tr><tr><td>T4</td><td>{M,U,C,K,Y}</td></tr><tr><td>T5</td><td>{C,O,O,K,I,E}</td></tr></table>		
	OR		
16	Discuss compact representation of frequent itemset (closed and maximal) with min-sup=3 (10M) <table><tr><td>Tid</td><td>items</td></tr><tr><td>T1</td><td>{A,B,C,D}</td></tr><tr><td>T2</td><td>{A,B,C,D}</td></tr><tr><td>T3</td><td>{A,B,C}</td></tr><tr><td>T4</td><td>{B,C,D}</td></tr><tr><td>T5</td><td>{C,D}</td></tr></table>	3	L4
17	What is Bayesian belief network ?explain with an example (10M)	4	L3
	OR		
18	Write KNN classification algorithm with an example. (10M)	4	L3
19	a) Write about evaluation of clustering algorithm? (5M) b) Explain about hierarchical clustering. (5M)	5	L2
	OR		
20	Discuss key issues involved in hierarchical and K-Means clustering. (10M)	5	L2