R20 Regulation *Subject code: 3P7FA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VII Semester Regular Examinations, November 2023

DATA WAREHOUSING AND DATA MINING

(Information Technology)

Maximum Marks: 70 Date:11.12.2023 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	Differentiate Data warehouse and DBMS.	L2
2	List and define the characteristics of Data warehouse.	L1
3	Write about data cleaning	L1
4	What is Data Discretization?	L1
5	What is the concept hierarchy generation.	L1
6	Define frequent sets, confidence and support	L1
7	What is meant by classification? What are the applications of classification?	L1
8	Write about Entropy and Information gain	L2
9	Differentiate OPTICS and DBSCAN.	L2
10	How are outliers detected using data mining?	L1
	Part-B	
Answer All the following questions.	(5X10M=50Marks)	
11	A. Discuss the star and snowflake schema in detail with suitable example. [5] B. Elaborate bind and listen functions in TCP sockets. [5]	L2 L2
	OR	
12	Define Data Cube computation. Explain the various methods for Data Cube Computation. Discuss Construction of Multi-dimensional model and its operations. [10]	L2
13	A. Explain the following with examples a) Discretization and binarization [3] b) Aggregation [2] B. List the steps of the Knowledge Discovery in Databases (KDD) and describe each of them. [5]	L2 L1
	OR	
14	A. List and define the measures of Similarity and Dissimilarity. [5] B. How to handle missing values in data sets? [5]	L1 L1

<table>
<tr><td>15</td><td>A database has four transactions. Let min_sup=60% and min_conf=80% [5+5]
<table><tr><th>TID</th><th>DATE</th><th>ITEMS_BROUGHT</th></tr><tr><td>100</td><td>10/15/99</td><td>{K,A,B,D}</td></tr><tr><td>200</td><td>10/15/99</td><td>{D,A,C,E,B}</td></tr><tr><td>300</td><td>10/19/99</td><td>{C,A,B,E}</td></tr><tr><td>400</td><td>10/22/99</td><td>{B,A,D}</td></tr></table>A) Find all frequent items using apriori & FP-growth, respectively. Compare the efficiency of the two meaning process.
B) List all of the strong association rules (with support s and confidence c) matching the following metarule where X is a variable representing customers, and item i denotes variables representing items (e.g., “A”, “B”, etc.): Vx Є transactions, buys (X,item1) ^ buys (X,item2) =>buys(X,item3)[s,c].</td><td>L3
L1</td></tr>
<tr><td></td><td>OR</td><td></td></tr>
<tr><td>16</td><td>A. Discuss the importance of Association Rule Mining. [5]
B. Write a note on Maximal Frequent Item Set and Closed Frequent Item Set. [5]</td><td>L2
L1</td></tr>
<tr><td>17</td><td>Write and explain about Classification by Back propagation Algorithm. [10]</td><td>L2</td></tr>
<tr><td></td><td>OR</td><td></td></tr>
<tr><td>18</td><td>Perform KNN classification for the data given below for X= {P1= 3, P2=7} where k=3 [10]
<table><tr><th>P1</th><th>P2</th><th>CLASS</th></tr><tr><td>7</td><td>7</td><td>FALSE</td></tr><tr><td>7</td><td>4</td><td>FALSE</td></tr><tr><td>3</td><td>4</td><td>TRUE</td></tr><tr><td>1</td><td>4</td><td>TRUE</td></tr></table></td><td>L3</td></tr>
<tr><td>19</td><td>A. Compare the performance of various outlier detection approaches. [5]
B. Write the key issue in hierarchical clustering algorithm. [5]</td><td>L3
L1</td></tr>
<tr><td></td><td>OR</td><td></td></tr>
<tr><td>20</td><td>A. What are the different clustering methods? Explain in detail. [5]
B. What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? [5]</td><td>L2
L2</td></tr>
</table>