



R20 Regulation

Subject code:3P6GB

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VI Semester Supplementary Examinations, December 2024

DATA WAREHOUSING AND DATA MINING (CSE (AIML))

Maximum Marks: 70

Date:30.12.2024

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	Differentiate between ROLAP and HOLAP.	1	L1
2	What is the use of metadata in data warehouse?	1	L1
3	List the methods of filling missing values.	2	L1
4	State the Apriori property.	2	L1
5	List the applications of data mining	3	L1
6	What is the need of confidence measure in association rule mining?	3	L1
7	Mention types of classifier techniques.	4	L1
8	Define Pre pruning and post pruning.	4	L1
9	Give a note on k - means algorithm.	5	L1
10	What are the weaknesses of hierarchical clustering?	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		CO	Bloom Tx
11	Compare and contrast online transaction processing with online analytical processing. [10M]	1	L3
OR			
12	Draw and explain the three tier data warehouse architecture. [10M]	1	L2
13	Describe the various phases in knowledge discovery process with a neat diagram. [10M]	2	L2
OR			
14	A) Write short notes on Measures Similarity and Dissimilarity. [4M] B) Explain data preprocessing in detail. [6M]	2	L2

15	Explain FP-Growth Algorithm and construct FP- Growth Tree.Support threshold=50%, Confidence= 60%, support threshold=50% and min_sup=3 for the given Table. [10M] <table><tr><td>Transaction</td><td>List of items</td></tr><tr><td>T1</td><td>I1,I2,I3</td></tr><tr><td>T2</td><td>I2,I3,I4</td></tr><tr><td>T3</td><td>I4,I5</td></tr><tr><td>T4</td><td>I1,I2,I4</td></tr><tr><td>T5</td><td>I1,I2,I3,I5</td></tr><tr><td>T6</td><td>I1,I2,I3,I4</td></tr></table>	Transaction	List of items	T1	I1,I2,I3	T2	I2,I3,I4	T3	I4,I5	T4	I1,I2,I4	T5	I1,I2,I3,I5	T6	I1,I2,I3,I4	3	L2
Transaction	List of items																
T1	I1,I2,I3																
T2	I2,I3,I4																
T3	I4,I5																
T4	I1,I2,I4																
T5	I1,I2,I3,I5																
T6	I1,I2,I3,I4																
	OR																
16	Apply apriori algorithm to find frequent itemsets from the following transactional database. Let min_sup = 30%. [10M] <table><tr><td>TID</td><td>Items_bought</td></tr><tr><td>1</td><td>Pen, notebook, ruler</td></tr><tr><td>2</td><td>Pencil, eraser, sharpener</td></tr><tr><td>3</td><td>Pen, ruler, chart, sharpener</td></tr><tr><td>4</td><td>Pencil, clip, eraser</td></tr><tr><td>5</td><td>Ruler, pin, story book, pen</td></tr><tr><td>6</td><td>Marker, chart, sketch pens</td></tr></table>	TID	Items_bought	1	Pen, notebook, ruler	2	Pencil, eraser, sharpener	3	Pen, ruler, chart, sharpener	4	Pencil, clip, eraser	5	Ruler, pin, story book, pen	6	Marker, chart, sketch pens	3	L2
TID	Items_bought																
1	Pen, notebook, ruler																
2	Pencil, eraser, sharpener																
3	Pen, ruler, chart, sharpener																
4	Pencil, clip, eraser																
5	Ruler, pin, story book, pen																
6	Marker, chart, sketch pens																
17	Describe the data classification process with a neat diagram. How does the NaiveBayesian classification works? Explain. [10M]	4	L2														
	OR																
18	A) State Bayes theorem How can it be applied for data classification? [5M] B) With example explain Bayesian belief network. [5M]	4	L2														
19	A)What are the advantages of PAM algorithm over k-means algorithm? [5M] B) Describe hierarchical clustering process. [5M]	5	L2														
	OR																
20	Discuss the similarity measures and distance measures frequently used in clustering the data. [10M]	5	L2														