

III YEAR B.TECH. EXAMINATION, DECEMBER -2022

Data Mining & Warehousing (900116)

Time: 3 Hrs.

Maximum Marks: 70

- Note: 1. Answer all five questions. All questions carry equal marks.
2. In each question part a, b, c are compulsory and part d has internal choice. Out of which part a & b carries 2 marks each, part c carries 3 marks and part d carries 7 marks.
3. All parts of each question are to be attempted at one place.
4. Assume suitable value for missing data, if any

Note:		1. Answer all five questions. All questions carry equal marks. 2. In each question part a, b, c are compulsory and part d has internal choice. Out of which part a & b carries 2 marks each, part c carries 3 marks and part d carries 7 marks. 3. All parts of each question are to be attempted at one place. 4. Assume suitable value for missing data, if any			
Q. No.			Marks	Course Outcomes	Bloom's Level
1.	(a)	What is data mining? How does KDD differ from data mining?	2	1	Remember
	(b)	What do you mean by Transactional databases? Discuss with example.	2	1,2	Understand
	(c)	Draw and discuss KDD (Knowledge Discovery in Databases) process model in brief.	3	1,2	Understand
	(d)	What are different ways of interfacing a data mining system with a database or data warehouse system?	7	1,2	Understand
		OR			
	(e)	Give the architecture of a data mining system. What are the essential components of a data mining system? Describe the purpose of each of these components.	7	1,2	Remember, Understand
2.	(a)	Define the term "Data Mart"	2	1,2	Understand
	(b)	What do you mean by the term 'Fact Tables'? Discuss in brief.	2	2,4	Understand
	(c)	Present an example where Data Warehousing is crucial to the success of a business. What Data Warehousing functions does this business need?	3	3,4	Understand, Apply
	(d)	What do you mean by data reduction techniques? Discuss attribute subset selection method with the help of suitable example.	7	2,3	Understand, Apply
		OR			
	(e)	Discuss about the following Data Warehousing schemas (i) Star Schema (ii) Snowflake Schema (iii) Fact Constellation Schema Also specify which schema is better and why.	7	2,3	Understand, Apply

3.	(a)	Discuss the need and importance of data preprocessing.	2	3	Remember										
	(b)	Discuss the need of Data Transformation with the help of suitable example.	2	3	Understand										
	(c)	Differentiate between Horizontal Data Layout and Vertical Data Layout.	3	3	Understand										
	(d)	Discuss the attribute subset selection method for Data Reduction.	7	3,4	Understand										
OR															
	(e)	Discuss about the following (i) Concept Hierarchy Generation (ii) Characterization	7	3	Understand										
4.	(a)	Define the term "Association Rule".	2	1,2	Understand										
	(b)	What do you understand by market basket analysis? Explain with the help of suitable example	2	4, 5	Apply										
	(c)	Explain the following (i) Frequent Items (ii) Support and Confidence	3	4, 5	Analyze, Apply										
	(d)	Explain the working of Partition Algorithm by considering suitable example dataset.	7	5	Analyze, Apply										
OR															
	(e)	A database has four transactions. Let the Minimum support is 50% <table border="1"><thead><tr><th>Tid</th><th>Items</th></tr></thead><tbody><tr><td>100</td><td>A,B,D,F</td></tr><tr><td>200</td><td>A,B,C,D,E</td></tr><tr><td>300</td><td>A,B,C,E</td></tr><tr><td>400</td><td>A,B,D</td></tr></tbody></table> Find all frequent items using Apriori algorithm.	Tid	Items	100	A,B,D,F	200	A,B,C,D,E	300	A,B,C,E	400	A,B,D	7	5	Evaluate
Tid	Items														
100	A,B,D,F														
200	A,B,C,D,E														
300	A,B,C,E														
400	A,B,D														
5.	(a)	What do you understand by the term cluster analysis?	2	2	Understand										
	(b)	Define the term "Classification".	2	3	Understand										
	(c)	Give the Name of major clustering methods available. Discuss any one of them in brief.	3	2,3	Remember, Understand										
	(d)	Consider any commercial data mining system and outline the major features of such a system	7	4, 5	Analyze										
OR															
	(e)	Discuss the various design issues for the efficient and intelligent data mining system.	7	4,5	Evaluate										
