



B.Tech VI Semester Regular/Supplementary Examinations, July 2024

DATA WAREHOUSING AND DATA MINING
 (CSE (AI & ML))

Maximum Marks: 70

Date: 22.07.2024 Duration: 3 hours

- Note:**
1. This question paper contains two parts A and B.
 2. Part A is compulsory which carries 10 marks. Answer all questions in Part A.
 3. Part B consists of 5 Units. Answer any one full question from each unit.
 4. Each question carries 10 marks and may have a, b, c, d as sub questions.

Part-A		CO	Bloom Tx
All the following questions carry equal marks (10X2M=20 Marks)			
1	Define Data warehouse.	CO1	L1
2	Explain Roll up operation in OLAP.	CO1	L1
3	Define Data mining.	CO2	L1
4	List out Data Transformation methods.	CO2	L1
5	List out the drawbacks of Apriori Algorithm.	CO3	L1
6	Write the applications of association analysis?	CO3	L1
7	Define decision tree.	CO4	L1
8	How prediction is different from classification?	CO4	L1
9	Write the strengths of hierarchical clustering?	CO5	L1
10	Compare agglomerative and divisive methods.	CO5	L1
Part-B			
Answer All the following questions. (5X10M=50Marks)			Bloom Tx level
11	Explain: "Star", "Snowflake" and "Fact constellation" schemas of data warehouse. [10]	CO1	L2
OR			
12	Clearly state the differences between "Data Warehouses" and "Operational Database Systems". [10]	CO1	L2
13	a) Discuss in detail about the steps of knowledge discovery. [5] b) Write a note on subset selection in attributes for data reduction. [5]	CO2	L2
OR			
14	a) Explain various data mining tasks. [5] b) Discuss briefly about data cleaning techniques. [5]	CO2	L2
15	State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set having minimum support value as 2 and minimum confidence value as 75%? [10]	CO3	L2

	TID	Items Purchased		
	T101	Cheese, milk, cookie		
	T102	Butter, milk, bread		
	T103	Cheese, butter, milk, bread		
	T104	Butter, bread		
	OR			
16	a) Write FP- growth algorithm. [5] b) Explain how association rules are generated from frequent item sets. [5]		CO3	L2
17	How does the Naïve Bayesian classification work? Explain in detail. [10]		CO4	L2
	OR			
18	Discuss K- Nearest Neighbor Classification-Algorithm and Characteristics. [10]		CO4	L2
19	a) What is outlier detection? Explain distance-based outlier detection. [5] b) Give a brief note on PAM Algorithm. [5]		CO5	L2
	OR			
20	a) Write K-means clustering algorithm. [5] b) Write the key issue in hierarchical clustering algorithm. [5]		CO5	L2