



## B.Tech IV Semester Regular Examinations, July 2024

### DATA WAREHOUSING & DATA MINING (CSE(DS))

**Maximum Marks: 60**

Date: 18.07.2024 Duration: 3 hours

- Note:**
1. This question paper contains two parts A and B.
  2. Part A is compulsory and 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 (10X1M=10 Marks)			
1. a)	Define Data warehousing.	CO1	L1
b)	List the Data warehouse Characteristics.	CO1	L1
c)	How can you go about filling in the missing values for this attribute?	CO2	L1
d)	Where we can apply the data reduction technique.	CO2	L1
e)	What are the challenges in data mining that motivate the mining tasks?	CO3	L1
f)	Define frequent item sets.	CO3	L1
g)	What is the use of attribute subset selection?	CO4	L1
h)	Mention different characteristics to construct a decision tree	CO4	L1
i)	Differentiate between clustering and classification.	CO5	L2
j)	How are outliers detected using data mining?	CO5	L1
Part-B			
Answer All the following questions. (5X10M=50Marks)			
2	Differentiate Operational database systems and data warehousing. Explain the star schema and fact constellation schemas. [10]	CO1	L3
OR			
3	a) Make a comparison between the MOLAP and HOLAP. [5] b) Discuss the star and snowflake schema in detail with a suitable example. [5]	CO1	L2 L6
4	Explain the various Data pre-processing techniques. How data reduction helps in data pre-processing. [10]	CO2	L2
OR			
5	a) What are the steps involved in KDD. [5] b) List and define the measures of Similarity and Dissimilarity. [5]	CO2	L1 L2
6	a) How can we mine multilevel Association rules efficiently using concept hierarchies? [3]	CO3	L1

	b) Illustrate with an A-priori algorithm for the given dataset with minsup=2. [7]		L2														
	<table border="1"> <thead> <tr> <th>TID</th> <th>List of items</th> </tr> </thead> <tbody> <tr> <td>001</td> <td>milk, dal, sugar, bread</td> </tr> <tr> <td>002</td> <td>Dal, sugar, wheat, jam</td> </tr> <tr> <td>003</td> <td>Milk, bread, curd, paneer</td> </tr> <tr> <td>004</td> <td>Wheat, paneer, dal, sugar</td> </tr> <tr> <td>005</td> <td>Milk, paneer, bread</td> </tr> <tr> <td>006</td> <td>Wheat, dal, paneer, bread</td> </tr> </tbody> </table>	TID	List of items	001	milk, dal, sugar, bread	002	Dal, sugar, wheat, jam	003	Milk, bread, curd, paneer	004	Wheat, paneer, dal, sugar	005	Milk, paneer, bread	006	Wheat, dal, paneer, bread		
TID	List of items																
001	milk, dal, sugar, bread																
002	Dal, sugar, wheat, jam																
003	Milk, bread, curd, paneer																
004	Wheat, paneer, dal, sugar																
005	Milk, paneer, bread																
006	Wheat, dal, paneer, bread																
	OR																
7	Can we design a FP-Growth method that mines the complete set of frequent item sets without candidate generation? If yes, explain with the example table mentioned above. [10]	CO3	L3														
8	Briefly explain the K-Nearest neighbor classification algorithm and its characteristics. [10]	CO4	L2														
	OR																
9	Describe the data classification process with a neat diagram. How does the Naïve Bayesian classification work? Explain. [10]	CO4	L6														
10	Explain the partitioning methods. Solve the following problem using Partition methods (K-means, K-medoids) for {2, 4, 10, 12, 8, 20, 30, 11, 25} where k = 2. [10]	CO5	L2														
	OR																
11	a) Differentiate Agglomerative and Divisive Hierarchical Clustering. Illustrate? [5] b) What are outliers? Discuss the methods adopted for outlier detection. [5]	CO5	L3 L6														