



**B.Tech IV Semester Supplementary Examinations, December 2024**

**DATA WAREHOUSING AND DATA MINING  
(CSE(DS))**

**Maximum Marks: 70**

**Date: 05.12.2024**

**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	What is Data Extraction?	1	L1
2	Define metadata.	1	L1
3	List the methods for filling missing values.	2	L1
4	Define the process of KDD.	2	L1
5	Mention the importance of association rule mining.	3	L1
6	State the property of Apriori.	3	L1
7	Write short note on entropy and information gain with formulae?	4	L1
8	List the measures for selecting best split in decision tree construction.	4	L1
9	Define clustering.	5	L1
10	List agglomerative and hierarchical clustering techniques.	5	L1

**Part-B**

Answer All the following questions. (5X10M=50Marks)		CO	Bloom Tx										
11	Draw and explain 3-tier architecture of a data warehouse. [10M]	1	L2										
OR													
12	List and explain OLAP operations with an example. [10M]	1	L2										
13	What is data mining? List and explain the motivating challenges of data mining. [10M]	2	L2										
OR													
14	With the help of a neat diagram Explain data mining as a step process of knowledge discovery. [10M]	2	L2										
15	Consider the following transactional data base , generate association rules and identify strong association rules using APRIORI algorithm with support=50% confidence=70% [10M]	3	L2										
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Tid</th> <th>Items</th> </tr> </thead> <tbody> <tr> <td>T1</td> <td>{1,3,4}</td> </tr> <tr> <td>T2</td> <td>{2,3,5}</td> </tr> <tr> <td>T3</td> <td>{1,2,3,5}</td> </tr> <tr> <td>T4</td> <td>{2,5}</td> </tr> </tbody> </table>	Tid	Items	T1	{1,3,4}	T2	{2,3,5}	T3	{1,2,3,5}	T4	{2,5}		
Tid	Items												
T1	{1,3,4}												
T2	{2,3,5}												
T3	{1,2,3,5}												
T4	{2,5}												

	OR																																							
16	Write and explain FP-growth algorithm with an example?	[10M]	3	L2																																				
17	Describe the general approaches to solve a classification problem.	[10M]	4	L2																																				
	OR																																							
18	Consider the following data, apply KNN algorithm to find class when acid durability=3 and strength=7	[10M]	4	L2																																				
	<table border="1"> <thead> <tr> <th>name</th> <th>acid durability</th> <th>strength</th> <th>class</th> </tr> </thead> <tbody> <tr> <td>Type1</td> <td>7</td> <td>7</td> <td>Bad</td> </tr> <tr> <td>Type2</td> <td>7</td> <td>4</td> <td>Bad</td> </tr> <tr> <td>Type3</td> <td>3</td> <td>4</td> <td>Good</td> </tr> <tr> <td>Type4</td> <td>1</td> <td>4</td> <td>Good</td> </tr> </tbody> </table>	name	acid durability	strength	class	Type1	7	7	Bad	Type2	7	4	Bad	Type3	3	4	Good	Type4	1	4	Good																			
name	acid durability	strength	class																																					
Type1	7	7	Bad																																					
Type2	7	4	Bad																																					
Type3	3	4	Good																																					
Type4	1	4	Good																																					
19	Explain agglomerative methods and construct a dendrogram for below distance matrix using complete linkage.	[10M]	5	L2																																				
	<table border="1"> <thead> <tr> <th></th> <th>1</th> <th>2</th> <th>3</th> <th>4</th> <th>5</th> </tr> </thead> <tbody> <tr> <th>1</th> <td>0</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>2</th> <td>9</td> <td>0</td> <td></td> <td></td> <td></td> </tr> <tr> <th>3</th> <td>3</td> <td>7</td> <td>0</td> <td></td> <td></td> </tr> <tr> <th>4</th> <td>6</td> <td>5</td> <td>9</td> <td>0</td> <td></td> </tr> <tr> <th>5</th> <td>11</td> <td>10</td> <td>2</td> <td>8</td> <td>0</td> </tr> </tbody> </table>		1	2	3	4	5	1	0					2	9	0				3	3	7	0			4	6	5	9	0		5	11	10	2	8	0			
	1	2	3	4	5																																			
1	0																																							
2	9	0																																						
3	3	7	0																																					
4	6	5	9	0																																				
5	11	10	2	8	0																																			
	OR																																							
20	Explain the outlier detection methods.	[10M]	5	L2																																				