



PRESIDENCY UNIVERSITY

BENGALURU

Roll No.														
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--

End - Term Examinations – MAY/ JUNE 2025

Date: 02-06-2025

Time: 01:00 pm – 04:00 pm

School: SOCSE	Program: B. Tech - CAI/COM/CSE	
Course Code: CSE2021	Course Name: Data Mining	
Semester: IV	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04	C05
Marks	14	14	24	24	24

Instructions:

- (i) Read all questions carefully and answer accordingly.
- (ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.

10Q x 2M=20M

1.	Define data mining.	2 Marks	L1	C01
2.	List the applications of data mining.	2 Marks	L1	C01
3.	Define Covariance.	2 Marks	L1	C02
4.	State the following (I)Object. (II) Attribute.	2 Marks	L1	C02
5.	Define Support.	2 Marks	L1	C03
6.	List two key differences between Apriori and FP-Growth algorithms.	2 Marks	L1	C03
7.	Name two types of unsupervised learning algorithms.	2 Marks	L1	C04
8.	Differentiate Supervised and Unsupervised Learning	2 Marks	L2	C04
9.	List examples of clustering applications.	2 Marks	L1	C05
10.	Define Pattern Mining.	2 Marks	L1	C05

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Explain knowledge discovery in database process with a neat diagram.	10 Marks	L2	CO1															
	b.	Describe data reduction in data mining in detail.	10 Marks	L2	CO2															
Or																				
12.	a.	Explain the Multidimensional view of data mining.	10 Marks	L2	CO1															
	b.	Determine dissimilarity by computing the Euclidean distance between the following pairs of points: <table><tr><td>Point</td><td>X</td><td>Y</td></tr><tr><td>p1</td><td>0</td><td>2</td></tr><tr><td>p2</td><td>2</td><td>0</td></tr><tr><td>p3</td><td>3</td><td>1</td></tr><tr><td>p4</td><td>5</td><td>1</td></tr></table>	Point	X	Y	p1	0	2	p2	2	0	p3	3	1	p4	5	1	10 Marks	L2	CO2
Point	X	Y																		
p1	0	2																		
p2	2	0																		
p3	3	1																		
p4	5	1																		

13.	a.	Discuss Decision Tree with all the Attribute Selection Measures.	5 Marks	L2	CO4										
	b	<div>Given the following transaction database, apply association rule mining using the Apriori algorithm. Identify all frequent itemsets with a minimum support count of 2</div> <table><tr><td>Tid</td><td>Items</td></tr><tr><td>10</td><td>A, C, D</td></tr><tr><td>20</td><td>B, C, E</td></tr><tr><td>30</td><td>A, B, C, E</td></tr><tr><td>40</td><td>B, E</td></tr></table>	Tid	Items	10	A, C, D	20	B, C, E	30	A, B, C, E	40	B, E	15 Marks	L3	CO3
Tid	Items														
10	A, C, D														
20	B, C, E														
30	A, B, C, E														
40	B, E														

Or

14.	a.	Describe the pseudocode for the Apriori algorithm used in Market Basket Analysis.	5 Marks	L2	C03																									
	b.	<div>Given a set of categorical weather conditions, predict whether tennis will be played or can go to cinema or can go to shopping or stay at home. The model will be trained using the Gini index decision tree algorithm. Calculate the root node.</div> <table><tr><td>Weekend</td><td>Weather</td><td>Parent</td><td>Money</td><td>Decision</td></tr><tr><td>1</td><td>Sunny</td><td>Yes</td><td>Rich</td><td>Cinema</td></tr><tr><td>2</td><td>Sunny</td><td>No</td><td>Rich</td><td>Tennis</td></tr><tr><td>3</td><td>Windy</td><td>Yes</td><td>Rich</td><td>Cinema</td></tr><tr><td>4</td><td>Rainy</td><td>Yes</td><td>Poor</td><td>Cinema</td></tr></table>	Weekend	Weather	Parent	Money	Decision	1	Sunny	Yes	Rich	Cinema	2	Sunny	No	Rich	Tennis	3	Windy	Yes	Rich	Cinema	4	Rainy	Yes	Poor	Cinema	15 Marks	L3	C04
Weekend	Weather	Parent	Money	Decision																										
1	Sunny	Yes	Rich	Cinema																										
2	Sunny	No	Rich	Tennis																										
3	Windy	Yes	Rich	Cinema																										
4	Rainy	Yes	Poor	Cinema																										

		5	Rainy	No	Rich	Stay in			
		6	Rainy	Yes	Poor	Cinema			
		7	Windy	No	Poor	Cinema			
		8	Windy	No	Rich	Shopping			
		9	Windy	Yes	Rich	Cinema			
		10	Sunny	No	Rich	Tennis			

15.	a.	Given the following transaction dataset, apply the FP-Growth algorithm to compute all frequent item sets with a minimum support count of 3. <div><table><tr><th>Transaction ID</th><th>Items</th></tr><tr><td>T1</td><td>{E, K, M, N, O, Y}</td></tr><tr><td>T2</td><td>{D, E, K, N, O, Y}</td></tr><tr><td>T3</td><td>{A, E, K, M}</td></tr><tr><td>T4</td><td>{C, K, M, U, Y}</td></tr><tr><td>T5</td><td>{C, E, I, K, O, O}</td></tr></table></div>	Transaction ID	Items	T1	{E, K, M, N, O, Y}	T2	{D, E, K, N, O, Y}	T3	{A, E, K, M}	T4	{C, K, M, U, Y}	T5	{C, E, I, K, O, O}	15 Marks	L3	C03
Transaction ID	Items																
T1	{E, K, M, N, O, Y}																
T2	{D, E, K, N, O, Y}																
T3	{A, E, K, M}																
T4	{C, K, M, U, Y}																
T5	{C, E, I, K, O, O}																
	b.	Describe Step-by-step Classification with a suitable diagram.	5 Marks	L2	C04												

Or

16.	a.	Explain FP growth tree in detail.	5 Marks	L2	C03																																													
	b.	<p>Consider the following dataset of weather conditions and corresponding target variable “Play”. Apply naïve base classifier to solve: If the weather is “Sunny” then the player should play or not.</p> <table><tr><th>SI No</th><th>Outlook</th><th>Play</th></tr><tr><td>0</td><td>Rainy</td><td>Yes</td></tr><tr><td>1</td><td>Sunny</td><td>Yes</td></tr><tr><td>2</td><td>Overcast</td><td>Yes</td></tr><tr><td>3</td><td>Overcast</td><td>Yes</td></tr><tr><td>4</td><td>Sunny</td><td>No</td></tr><tr><td>5</td><td>Rainy</td><td>Yes</td></tr><tr><td>6</td><td>Sunny</td><td>Yes</td></tr><tr><td>7</td><td>Overcast</td><td>Yes</td></tr><tr><td>8</td><td>Rainy</td><td>No</td></tr><tr><td>9</td><td>Sunny</td><td>No</td></tr><tr><td>10</td><td>Sunny</td><td>Yes</td></tr><tr><td>11</td><td>Rainy</td><td>No</td></tr><tr><td>12</td><td>Overcast</td><td>Yes</td></tr><tr><td>13</td><td>Overcast</td><td>Yes</td></tr></table>	SI No	Outlook	Play	0	Rainy	Yes	1	Sunny	Yes	2	Overcast	Yes	3	Overcast	Yes	4	Sunny	No	5	Rainy	Yes	6	Sunny	Yes	7	Overcast	Yes	8	Rainy	No	9	Sunny	No	10	Sunny	Yes	11	Rainy	No	12	Overcast	Yes	13	Overcast	Yes	15 Marks	L3	C04
SI No	Outlook	Play																																																
0	Rainy	Yes																																																
1	Sunny	Yes																																																
2	Overcast	Yes																																																
3	Overcast	Yes																																																
4	Sunny	No																																																
5	Rainy	Yes																																																
6	Sunny	Yes																																																
7	Overcast	Yes																																																
8	Rainy	No																																																
9	Sunny	No																																																
10	Sunny	Yes																																																
11	Rainy	No																																																
12	Overcast	Yes																																																
13	Overcast	Yes																																																

17.	a.	Apply K-means clustering algorithm to divide the following data into two clusters[K=2]. <table><tr><td>X1</td><td>1</td><td>2</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>X2</td><td>1</td><td>1</td><td>3</td><td>2</td><td>3</td><td>5</td></tr></table>	X1	1	2	2	3	4	5	X2	1	1	3	2	3	5	10 Marks	L3	C05																																			
X1	1	2	2	3	4	5																																																
X2	1	1	3	2	3	5																																																
	b.	Apply your understanding of clustering by explaining different clustering approaches (such as partitioning, hierarchical, and density-based methods) with suitable examples for each.	10 Marks	L3	C05																																																	
Or																																																						
18.	a.	Consider the following data matrix, apply single linkage in agglomerative clustering method and solve with forming the clusters and dendrograms. <table><tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td></tr><tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>B</td><td>5</td><td>0</td><td></td><td></td><td></td><td></td></tr><tr><td>C</td><td>14</td><td>9</td><td>0</td><td></td><td></td><td></td></tr><tr><td>D</td><td>11</td><td>20</td><td>13</td><td>0</td><td></td><td></td></tr><tr><td>E</td><td>18</td><td>15</td><td>6</td><td>3</td><td>0</td><td></td></tr><tr><td>F</td><td>10</td><td>16</td><td>8</td><td>10</td><td>11</td><td></td></tr></table>		A	B	C	D	E	F	A	0						B	5	0					C	14	9	0				D	11	20	13	0			E	18	15	6	3	0		F	10	16	8	10	11		10 Marks	L3	C05
	A	B	C	D	E	F																																																
A	0																																																					
B	5	0																																																				
C	14	9	0																																																			
D	11	20	13	0																																																		
E	18	15	6	3	0																																																	
F	10	16	8	10	11																																																	
	b.	Compare agglomerative and divisive hierarchical clustering methods. Interpret how linkage methods (single, complete, average) affect cluster formation with examples.	10 Marks	L3	C05																																																	