



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

# PRESIDENCY UNIVERSITY

## BENGALURU

### End - Term Examinations - DECEMBER 2025

Date: 15-12-2025

Time: 1.00pm to 04.00pm

<b>School:</b> SOCSE	<b>Program:</b> B.TECH- Computer Science and Engineering (CSD)		
<b>Course Code :</b> CSE3512	<b>Course Name :</b> DATA MINING		
<b>Semester:</b> V	<b>Max Marks:</b> 100	<b>Weightage:</b> 50%	

CO - Levels	C01	C02	C03	C04	C05
Marks					

#### 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 x2M=20M

1	Define data mining ?	2 Marks	L1	C01
2	Define the term Data Cleaning with example.	2 Marks	L1	C01
3	Explain clustering and regression with example	2 Marks	L2	C02
4	What is Z-Score normalization?	2 Marks	L1	C02
5	Describe the capabilities of OLAP?	2 Marks	L1	C03
6	Describe Bayesian Classification?	2 Marks	L1	C03
7	Explain Data Types in Clustering?	2 Marks	L2	C04
8	Explain definition of Apriori algorithm?	2 Marks	L2	C04
9	Explain the various types of Data?	2 Marks	L2	C05
10	Derive Outlier analysis	2 Marks	L1	C05

**Part B**

**Answer the Questions**

**Total 80 Marks**

<b>11.</b>	<b>a.</b>	Elucidate Data mining Functionalities with suitable examples?	<b>10 Marks</b>	<b>L2</b>	<b>C01</b>
------------	-----------	---	---------------------	-----------	------------

**Or**

<b>12.</b>	<b>b.</b>	Explain the Various Issues in data mining with example?	<b>10 Marks</b>	<b>L2</b>	<b>C01</b>
------------	-----------	---	---------------------	-----------	------------

<b>13.</b>	<b>a.</b>	Explain the different types of attributes used in data mining. Provide examples for each type and discuss how these types influence the choice of data mining techniques.	<b>10 Marks</b>	<b>L2</b>	<b>C02</b>
------------	-----------	---	---------------------	-----------	------------

**Or**

<b>14.</b>	<b>b.</b>	<p>Explain about Normalization in detail . Suppose that the minimum and maximum values for the attribute income are \$12,000 and \$98,000, respectively. We would like to map income to the range [0.0,1.0]. By min-max normalization, a value of \$73,600 for income is transformed find the normalization. Apply</p> <p>i. Min-Max ii. Z-Score iii. Decimal Scaling</p>	<b>10 Marks</b>	<b>L3</b>	<b>C02</b>
------------	-----------	---	---------------------	-----------	------------

<b>15.</b>	<b>a.</b>	Explain about Rule based classification in detail with suitable example .	<b>10 Marks</b>	<b>L2</b>	<b>C04</b>
------------	-----------	---	---------------------	-----------	------------

**Or**

<b>16.</b>	<b>b.</b>	<p>For the given data set apply the classification using Naïve Bayes for the data Red, Domestic, SUV.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Color</th> <th>Type</th> <th>Origin</th> <th>stolen</th> </tr> </thead> <tbody> <tr><td>Red</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr> <tr><td>Red</td><td>Sports</td><td>Domestic</td><td>No</td></tr> <tr><td>Red</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr> <tr><td>Yellow</td><td>Sports</td><td>Domestic</td><td>No</td></tr> <tr><td>Yellow</td><td>Sports</td><td>Imported</td><td>Yes</td></tr> <tr><td>Yellow</td><td>SUV</td><td>Imported</td><td>No</td></tr> <tr><td>Yellow</td><td>SUV</td><td>Imported</td><td>Yes</td></tr> <tr><td>Yellow</td><td>SUV</td><td>Domestic</td><td>No</td></tr> <tr><td>Red</td><td>SUV</td><td>Imported</td><td>No</td></tr> <tr><td>Red</td><td>Sports</td><td>Imported</td><td>Yes</td></tr> </tbody> </table>	Color	Type	Origin	stolen	Red	Sports	Domestic	Yes	Red	Sports	Domestic	No	Red	Sports	Domestic	Yes	Yellow	Sports	Domestic	No	Yellow	Sports	Imported	Yes	Yellow	SUV	Imported	No	Yellow	SUV	Imported	Yes	Yellow	SUV	Domestic	No	Red	SUV	Imported	No	Red	Sports	Imported	Yes	<b>10 Marks</b>	<b>L3</b>	<b>C04</b>
Color	Type	Origin	stolen																																														
Red	Sports	Domestic	Yes																																														
Red	Sports	Domestic	No																																														
Red	Sports	Domestic	Yes																																														
Yellow	Sports	Domestic	No																																														
Yellow	Sports	Imported	Yes																																														
Yellow	SUV	Imported	No																																														
Yellow	SUV	Imported	Yes																																														
Yellow	SUV	Domestic	No																																														
Red	SUV	Imported	No																																														
Red	Sports	Imported	Yes																																														

<b>17.</b>  <b>a.</b>	For the given transactional data find the interesting pattern. Use Apriori algorithm have min sup= 2 and min -conf= 50%. Find the association and efficiency of the algorithm.			<b>15</b>  <b>Marks</b>	<b>L3</b>	<b>C03</b>
	<b>Transaction ID</b>	<b>Items Purchased</b>				
	1	Bread, Cheese, Egg, Juice				
	2	Bread, Cheese, Juice				
	3	Bread, Milk, Yogurt				
	4	Bread, Juice, Milk				
5	Cheese, Juice, Milk					

**Or**

<b>18.</b>  <b>b.</b>	What do you understand by Classification by Backpropagation? Solve the given network using initial values. Find the new weights of the network.			<b>15</b>  <b>Marks</b>	<b>L3</b>	<b>C04</b>																								
	<p>Initial Input, Weight, and Bias Values</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th><math>x_1</math></th> <th><math>x_2</math></th> <th><math>x_3</math></th> <th><math>w_{14}</math></th> <th><math>w_{15}</math></th> <th><math>w_{24}</math></th> <th><math>w_{25}</math></th> <th><math>w_{34}</math></th> <th><math>w_{35}</math></th> <th><math>w_{46}</math></th> <th><math>w_{56}</math></th> <th><math>\theta_4</math></th> <th><math>\theta_5</math></th> <th><math>\theta_6</math></th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0</td> <td>1</td> <td>0.2</td> <td>-0.3</td> <td>0.4</td> <td>0.1</td> <td>-0.5</td> <td>0.2</td> <td>-0.3</td> <td>-0.2</td> <td>-0.4</td> <td>0.2</td> <td>0.1</td> </tr> </tbody> </table>						$x_1$	$x_2$	$x_3$	$w_{14}$	$w_{15}$	$w_{24}$	$w_{25}$	$w_{34}$	$w_{35}$	$w_{46}$	$w_{56}$	$\theta_4$	$\theta_5$	$\theta_6$	1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3
$x_1$	$x_2$	$x_3$	$w_{14}$	$w_{15}$	$w_{24}$	$w_{25}$	$w_{34}$	$w_{35}$	$w_{46}$	$w_{56}$	$\theta_4$	$\theta_5$	$\theta_6$																	
1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2	-0.4	0.2	0.1																	

<b>19.</b>  <b>a.</b>	Apply entropy based discretization on the given set $S = (16,n),(0,y),(4,y),(12,y),(16,n),(26,n),(18,y),(24,n),(28,n)$ . If $S$ has partitioned in to 2 intervals $S_1$ & $S_2$ with 2 possible split points 14 & 21. Find the Best split point.			<b>15</b>  <b>Marks</b>	<b>L2</b>	<b>C04</b>

**Or**

<b>20.</b>  <b>b.</b>	Explain about Partationing clustering with algorithm and solve the problem by using ecludian distance measure: Define it with neat visual representations. Suppose we want to group the visitors to a website using just their age as follows:			<b>15</b>  <b>Marks</b>	<b>L2</b>	<b>C05</b>

S.No	X	Y
1	16	16
2	20	20
3	21	22
4	29	36
5	42	43
6	45	61
7	66	54

21.	a.	A database has five transactions. Let min sup = 3 and min confidence= 80%	20 Marks	L3	CO3												
		<table border="1"> <thead> <tr> <th>Transaction ID</th> <th>List of Item sets</th> </tr> </thead> <tbody> <tr> <td>T100</td> <td>A,B</td> </tr> <tr> <td>T200</td> <td>A,C,D</td> </tr> <tr> <td>T300</td> <td>A,B,C,D</td> </tr> <tr> <td>T400</td> <td>A,D,E</td> </tr> <tr> <td>T500</td> <td>B,C</td> </tr> </tbody> </table>				Transaction ID	List of Item sets	T100	A,B	T200	A,C,D	T300	A,B,C,D	T400	A,D,E	T500	B,C
		Transaction ID				List of Item sets											
		T100				A,B											
		T200				A,C,D											
		T300				A,B,C,D											
		T400				A,D,E											
T500	B,C																
Apply FP-growth tree and find Conditional Pattern Base and Conditional FPTree.																	

**Or**

22.	B.	Use the distance matrix to perform single link and complete link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.	20 Marks	L3	CO5																																				
		<table border="1"> <thead> <tr> <th></th> <th>P1</th> <th>P2</th> <th>P3</th> <th>P4</th> <th>P5</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>0.00</td> <td>0.10</td> <td>0.41</td> <td>0.55</td> <td>0.35</td> </tr> <tr> <td>P2</td> <td>0.10</td> <td>0.00</td> <td>0.64</td> <td>0.47</td> <td>0.98</td> </tr> <tr> <td>P3</td> <td>0.41</td> <td>0.64</td> <td>0.00</td> <td>0.44</td> <td>0.85</td> </tr> <tr> <td>P4</td> <td>0.55</td> <td>0.47</td> <td>0.44</td> <td>0.00</td> <td>0.76</td> </tr> <tr> <td>P5</td> <td>0.35</td> <td>0.98</td> <td>0.85</td> <td>0.76</td> <td>0.00</td> </tr> </tbody> </table>					P1	P2	P3	P4	P5	P1	0.00	0.10	0.41	0.55	0.35	P2	0.10	0.00	0.64	0.47	0.98	P3	0.41	0.64	0.00	0.44	0.85	P4	0.55	0.47	0.44	0.00	0.76	P5	0.35	0.98	0.85	0.76	0.00
						P1	P2	P3	P4	P5																															
		P1				0.00	0.10	0.41	0.55	0.35																															
		P2				0.10	0.00	0.64	0.47	0.98																															
		P3				0.41	0.64	0.00	0.44	0.85																															
		P4				0.55	0.47	0.44	0.00	0.76																															
P5	0.35	0.98	0.85	0.76	0.00																																				