



PRESIDENCY UNIVERSITY

BENGALURU

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

End - Term Examinations - December 2025

Date: 11 - 12- 2025

Time: 09:30am - 12:30pm

School: SOIS	Program: BCA		
Course Code: CSA1703	Course Name: Data Mining		
Semester: III	Max Marks: 100	Weightage: 50%	

CO - Levels	C01	C02	C03	C04	C05
Marks	24	26	26	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.	State Data Mining and list any two applications of it.	2 Marks	L1	C01
2.	Recall whether the Data mining is misnomer? Justify.	2 Marks	L1	C01
3.	List the types of attributes.	2 Marks	L1	C02
4.	Let p, q are two binary valued attributes. The values of p & q are specified below $p=100000000$, $q=0000001001$ Compute SMC using p&q values.	2 Marks	L2	C02
5.	Distinguish between Pure bin and Impure bin.	2 Marks	L1	C02
6.	State the 2 differences between FP growth and Apriori method.	2 Marks	L1	C03
7.	A dataset has 7 transactions, and the itemset {A,B} appears in 3 transactions. Compute the support for {A,B}?	2 Marks	L2	C03
8.	Tell the difference between Supervised and Unsupervised learning.	2 Marks	L1	C03
9.	Differentiate between intra-cluster and inter-cluster with an example.	2 Marks	L2	C04
10.	Define Outlier with a suitable example.	2 Marks	L1	C04

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Explain the various stages involved in the Data Mining process with the help of a neat diagram.	10 Marks	L2	CO1
	b.	Enumerate various Data Mining functionalities with a suitable example.	10 Marks	L2	CO1
Or					
12.	a.	Draw a typical view from ML and Statistics. Portray.	10 Marks	L2	CO1
	b.	Illustrate major issues of data mining regarding mining methodology, user interactions, Efficiency and scalability ,diversity of data types and Society	10 Marks	L2	CO1

13.	a.	<p>A smart agriculture system records the soil moisture levels (in millivolts) from IoT sensors installed in different crop fields. The collected data values are:</p> <p style="padding-left: 20px;">200, 400, 500, 600, 800</p> <p>Apply Min-Max, Z-Score and decimal scaling normalization to the given data set.</p>	10 Marks	L3	CO2
	b.	<p>Rachel told Eric that the reason her car insurance is less expensive is that female drivers get in fewer accidents than male drivers. Specifically, she says that male drivers are held responsible in 65% of accidents involving drivers under 23. If Eric does some research of his own and discovers that 46 out of the 85 accidents he investigates involve male drivers, Compute his data support or refute Rachel's hypothesis? Assume level of significance as 0.05.</p> <p>Note: $DF(1,0.05)=3.841$; $DF(2,0.05)=5.991$.</p>	10 Marks	L3	CO2

Or

14.	a.	<p>Given a set of samples $S=(60,N),(75,N),(70,N),(90,Y),(85,Y),(95,Y),(100,N),(120,N),(125,N), (220,N)$. If S has to be partitioned in 2 intervals S1 & S2 for the split points 80 & 97. Apply Entropy based Discretization to determine the Best Split Point.</p>	10 Marks	L3	CO2
	b.	<p>From the following data 11,13,13,15,15,16,19,20,20,20,21,21,45,45,71,72,73,75. Apply 3 different binning methods for smoothing the data. Each bin consisting of 6 values.</p>	10 Marks	L3	CO2

15.	a.	<p>A database has 5 transactions. Apply FP Growth algorithm to:</p> <p>i) Generate the frequent itemset ii) Construct FP tree iii) Find conditional frequent pattern and frequent pattern generation. Let the minimum support be 3.</p> <table border="1"> <thead> <tr> <th>TID</th> <th>ITEMS</th> </tr> </thead> <tbody> <tr> <td>T1</td> <td>{M,O,N,K,E,Y}</td> </tr> <tr> <td>T2</td> <td>{D,O,N,K,E,Y}</td> </tr> <tr> <td>T3</td> <td>{M,A,K,E}</td> </tr> <tr> <td>T4</td> <td>{M,U,C,K,Y}</td> </tr> <tr> <td>T5</td> <td>{C,O,K,I,E}</td> </tr> </tbody> </table>	TID	ITEMS	T1	{M,O,N,K,E,Y}	T2	{D,O,N,K,E,Y}	T3	{M,A,K,E}	T4	{M,U,C,K,Y}	T5	{C,O,K,I,E}	20 Marks	L3	CO3
TID	ITEMS																
T1	{M,O,N,K,E,Y}																
T2	{D,O,N,K,E,Y}																
T3	{M,A,K,E}																
T4	{M,U,C,K,Y}																
T5	{C,O,K,I,E}																

Or

16.	a.	<p>Apply ID3 Algorithm to find the root node in a decision tree using the following data set:</p> <table border="1"> <thead> <tr> <th>age</th> <th>income</th> <th>student</th> <th>credit_rating</th> <th>buys_computer</th> </tr> </thead> <tbody> <tr><td><=30</td><td>high</td><td>no</td><td>fair</td><td>no</td></tr> <tr><td><=30</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr> <tr><td>31...40</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr> <tr><td>>40</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr> <tr><td>>40</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>>40</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr> <tr><td>31...40</td><td>low</td><td>yes</td><td>excellent</td><td>yes</td></tr> <tr><td><=30</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr> <tr><td><=30</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>>40</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td><=30</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr> <tr><td>31...40</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr> <tr><td>31...40</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr> <tr><td>>40</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr> </tbody> </table>	age	income	student	credit_rating	buys_computer	<=30	high	no	fair	no	<=30	high	no	excellent	no	31...40	high	no	fair	yes	>40	medium	no	fair	yes	>40	low	yes	fair	yes	>40	low	yes	excellent	no	31...40	low	yes	excellent	yes	<=30	medium	no	fair	no	<=30	low	yes	fair	yes	>40	medium	yes	fair	yes	<=30	medium	yes	excellent	yes	31...40	medium	no	excellent	yes	31...40	high	yes	fair	yes	>40	medium	no	excellent	no	20 Marks	L3	CO3
age	income	student	credit_rating	buys_computer																																																																												
<=30	high	no	fair	no																																																																												
<=30	high	no	excellent	no																																																																												
31...40	high	no	fair	yes																																																																												
>40	medium	no	fair	yes																																																																												
>40	low	yes	fair	yes																																																																												
>40	low	yes	excellent	no																																																																												
31...40	low	yes	excellent	yes																																																																												
<=30	medium	no	fair	no																																																																												
<=30	low	yes	fair	yes																																																																												
>40	medium	yes	fair	yes																																																																												
<=30	medium	yes	excellent	yes																																																																												
31...40	medium	no	excellent	yes																																																																												
31...40	high	yes	fair	yes																																																																												
>40	medium	no	excellent	no																																																																												

17.	a.	<p>Illustrate the key challenges you might encounter when detecting anomalies.</p>	10 Marks	L3	CO4
	b.	<p>A delivery company wants to optimize its regional warehouse locations. They have collected the coordinates of delivery hotspots in a city as follows:</p> <p>(2,3), (5,6), (8,7), (1,4), (2,2), (6,7), (3,4), (8,6)</p> <p>Apply K-Means clustering to group these delivery hotspots into 2 clusters. Assume C1=(2,3) & C2=(5,6)</p>	10 Marks	L3	CO4

Or

18.**a.**

A healthcare organization wants to cluster six hospitals (A, B, C, D, E, and F) based on their patient treatment similarity index. The pairwise distance matrix between the hospitals is given. Apply Complete Linkage Agglomerative Hierarchical Clustering and illustrate the hierarchy using a dendrogram.

	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	0

20 Marks**L3****CO4**