PRESIDENCY UNIVERSITY

BENGALURU

Make-up Examinations – December 2025	
Date: 29- 12- 2025	**Time:** 01:00pm – 04:00pm

School: SOIS	Program: BCA	
Course Code : CSA2007	**Course Name :** Data Mining	
Semester: MK	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	24	26	24	26	

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 KDD	2 Marks	L1	CO1
2	Outline the 2 disadvantages of Data Mining.	2 Marks	L1	CO1
3	Let p, q are two binary valued attributes. The values of p & q are specified below p= 1 0 0 0 0 0 0 0 0 0, q = 0 0 0 0 0 0 1 0 0 1 Compute SMC using p&q values.	2 Marks	L2	CO2
4	List the various reduction techniques in data preprocessing.	2 Marks	L1	CO2
5	Describe the relationship between covariance and correlation coefficient.	2 Marks	L2	CO2
6	Recite the benefits of FP growth over Apriori method.	2 Marks	L1	CO3
7	Differentiate between Support and Confidence.	2 Marks	L2	CO3
8	What is Lazy learner and Eager learner ? Give example.	2 Marks	L2	CO4
9	List the difference between Supervised and Unsupervised learning.	2 Marks	L1	CO4
10	Define Outlier with a suitable example.	2 Marks	L1	CO4

Part B

Answer the Questions   **Total 80 Marks.**

11.	a.	Elucidate Data mining Functionalities with an example.	10 Marks	L2	CO1
	b.	Explain the various major issues and challenges associated in Data Mining.	10 Marks	L2	CO1

Or

12.	a.	Describe the steps involved in knowledge discovery process with a neat diagram.	10 Marks	L2	CO1
	b.	Explain the role of data mining in business intelligence with a suitable diagram.	10 Marks	L2	CO1

13.	a.	From the following data 11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75. Apply 3 different binning methods for smoothing the data.	10 Marks	L3	CO2
	b.	Compute Z-Score and decimal scaling normalization for the given data : 200,400,500,600,800,1000	10 Marks	L2	CO2

Or

14.	a.	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. Compute the Best Split Point.	10 Marks	L2	CO2
	b.	Explain the different Attribute subset selection methods.	10 Marks	L2	CO2

15.	a.	Determine association rules using Apriori Algorithm for the given TDB. Let assume min support=3 and min confidence=50%	20 Marks	L3	CO3

Tid	Items
T1	I1,I3,I7
T2	I2,I3,I7
T3	I1,I2,I3
T4	I2,I3
T5	I2,I3,I4,I5
T6	I2,I3
T7	I1,I2,I3,I4,I6
T8	I2,I3,I4,I6
T9	I1
T10	I1,I3

Or

| 16. | a. | Apply FP-Growth algorithm to generate association rules for the given data set.

TID   **Items bought**
100   **{f, a, c, d, g, i, m, p}**
200   **{a, b, c, f, l, m, o}**
300   **{b, f, h, j, o}**
400   **{b, c, k, s, p}**
500   **{a, f, c, e, l, p, m, n}** | 20 Marks | L3 | CO3 |

| 17. | a. | Predict Buys_Computer, when (age <= 30 , income = medium, student = yes, credit_rating = fair) from the given table using Naives Bayesian algorithm. | 10 Marks | L3 | CO4 |

Age	Income	Student	Credit_Rating	Buys_Computer
Less than 30	High	No	Fair	No
Less than 30	High	No	Excellent	No
Between 31 to 40	High	No	Fair	Yes
Greater than 40	Medium	No	Fair	Yes
Greater than 40	Low	Yes	Fair	Yes
Greater than 40	Low	Yes	Excellent	No
Between 31 to 40	Low	Yes	Excellent	Yes
Less than 30	Medium	No	Fair	No
Less than 30	Low	Yes	Fair	Yes
Greater than 40	Medium	Yes	Fair	Yes
Less than 30	Medium	Yes	Excellent	Yes
Between 31 to 40	Medium	No	Excellent	Yes
Between 31 to 40	High	Yes	Fair	Yes
Greater than 40	Medium	No	Excellent	No

| | b. | Using the provided dataset and the KNN algorithm with k=3, predict whether the new employee should be classified as a Good or Bad performer. | 10 Marks | L3 | CO4 |

X1	X2	Performance
40	50	Bad
95	75	Good
60	70	Good
45	35	Bad
80	85	Good

Or

| 18. | a. | A real estate company wants to classify six cities (A, B, C, D, E, and F) into clusters based on their proximity to each other. The company uses the following distance matrix (in kilometers) between cities to group them hierarchically, starting with each city as its own cluster. Use Single and Complete Linkage | 20 Marks | L3 | CO4 |

agglomerative HCA, also draw dendrogram to represent the hierarchy of clusters.

	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

***** BEST WISHES *****