Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 08-10-2025 **Time:** 09.30am to 11.00am

School: SOIS	**Program:** BCA	
Course Code : CSA1703	**Course Name:** Data Mining	
Semester: III	**Max Marks**: 50	**Weightage**: 25%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**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. **5Q x 2M=10M**

1	Outline 4 applications of Data Mining.	2 Marks	L1	CO1
2	Recall the four alternative names of Data Mining.	2 Marks	L1	CO1
3	p = 1 0 0 0 1 0 0 0 1 0 q = 0 1 0 0 0 0 1 0 0 1 Apply SMC method for the given data.	2 Marks	L3	CO2
4	State the common properties of similarity.	2 Marks	L1	CO2
5	Given the dataset divide it into 3 bins and apply Smoothing by Bin Boundaries for the given data set. [3,5,7,8,12,18,25,30,45]	2 Marks	L3	CO2

Part B

Answer the Questions. **Total Marks 40M**

6.	a.	Elaborate the stages of the Data Mining process with a suitable diagram.	10 Marks	L2	CO1
	b.	Elucidate the 3 advantages and role of data mining in business intelligence with a neat diagram.	10 Marks	L2	CO1
		Or			
7.	a.	Explain functionalities of data mining in detail with suitable example.	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
8.	a.	Apply data cleaning techniques (such as handling missing values and noise removal) and data integration techniques (such as schema integration, entity identification, and derivable data) on a sample data, and demonstrate how the quality of data improves after processing.	10 Marks	L3	CO2
	b.	Rakhi 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 under23. If Eric does some research of his own and discovers that 46 out of the 85 accidents, he investigates involve male drivers, does his data support or refute Rachel's hypothesis? Apply Chi-square test. Assume that 0.05 level of significance. Assume that H_0:Male drivers are responsible for 65% of accidents and female drivers are responsible for 35%. H_1:The data do not match the proposed model Note: DF=(1,0.05)= 3.841 ; DF(1,0.01)= 6.635 DF=(2,0.05)= 5.991 ; DF(2,0.01)= 9.210	10 Marks	L3	CO2
		Or			
9.	a.	Demonstrate the types of attributes and their properties with suitable example.	10 Marks	L3	CO2
	b.	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. Determine the Best Split Point.	10 Marks	L3	CO2