PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 07-10-2025   **Time:** 02.00pm to 03.30pm

School: SOCSE	**Program:** B.Tech. Computer Science and Engineering- CBD
Course Code :CBD2506	**Course Name:** Data Mining And Predictive Analysis
Semester:V	**Max Marks:**50 **Weightage:**25%

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

1	Define Data Mining?	2 Marks	L1	CO1
2	What is data cleaning? Give one example.	2 Marks	L1	CO1
3	What is the limitation of logistic regression.	2 Marks	L1	CO2
4	Distinguish between classification and prediction.	2 Marks	L4	CO2
5	Define Support Vector Machine (SVM).	2 Marks	L1	CO2

Part B

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

6.	a.	Explain different forms of knowledge that can be mined from databases with suitable examples.	10 Marks	L2	CO 1
		Or			
7.	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. Demonstarte the Best SplitPoint.	10 Marks	L2	CO 1

8.	a.	Define Normalization, Utilise the data for attribute age answer the following: (a) Use min-max normalization to transform the value 35 for age onto the min and max of 25 & 45 with the of range [0.0,1.0]. (b) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years and mean are 10.55. (c) Use normalization by decimal scaling to transform the value 35 for age.	10 Marks	L3	CO 1
		Or			
9.	a.	Define Data cleaning in detail and apply Binning on the given data . S = { 13, 15, 9, 16, 10, 20, 21, 22, 22, 25 } .	10 Marks	L3	CO 1

10.	a.	Distinguish between classification and prediction with examples.	10 Marks	L4	CO 2
		Or			
11.	a.	Expalin Navie Bayes , By utilizing a naive Bayes classification Find the probability outcome of derived data.	10 Marks	L3	CO 2

Color	Legs	Height	Smelly	Species
White	3	Short	Yes	M
Green	2	Tall	No	M
Green	3	Short	Yes	M
White	3	Short	Yes	M
Green	2	Short	No	H
White	2	Tall	No	H
White	2	Tall	No	H
White	2	Short	Yes	H

12.	a.	What are the multidimensional classification alogithms , Solve using K-nearest neighbor to form clusters, where centroids are A3,A4.	10 Marks	L3	CO 2

Point	Coordinates
A1	(2,10)
A2	(2, 6)
A3	(11,11)
A4	(6, 9)
A5	(6, 5)
A6	(1, 2)
A7	(5, 10)

Or

13.	a.	*Demonstarte Logistic regression in detail and their techniques.*	10 Marks	L2	CO 2