Roll No: | | | | | | | | | | | |

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 07/10/2026 **Time:** 02:00 PM - 03:30 PM

Course Code: CBD2506 **Course Name:** Data Mining and Predictive Analytics

Semester: Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	44	46

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	**List any two effects of poor data quality on data mining results.**	2	CO1	BT2
2	**Apply data cleaning and transformation techniques on the dataset {'10', '20', NULL, 'thirty'} to handle missing values and inconsistent formats.**	2	CO1	BT2
3	**What is the purpose of the ROC Curve in evaluating classifiers?**	2	CO2	BT2
4	**A classifier has 30 true positives (TP) and 10 false negatives (FN). Compute the Recall and briefly state what it indicates.**	2	CO2	BT3
5	**What is the posterior score in Bayesian classification?**	2	CO2	BT2

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	**A bank wants to detect fraudulent transactions from its transaction dataset. Apply the steps of the KDD process to achieve this goal.**	10	CO1	BT3
7	Apply your knowledge of different types of data and data formats to classify a given set of data with suitable examples.	10	CO1	BT3

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	Apply different data integration approaches to combine data from multiple sources.	10	CO1	BT3
9	Illustrate the different pattern types with suitable data examples.	10	CO1	BT3

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT

10	**The dataset of PASS or FAIL in an exam of 5 students is given in the table. Use logistic regression as classifier to answer the following questions.** 1. **Calculate the probability of pass for the student who studied 33 hours.** 2. **At least how many hours student should study that makes he will pass the course with the probability of more than 95%.** **Assume the model suggested by the optimizer for odds of passing the course is, log(odds) = -64 +2 * hours** **Hours Study Pass (1) / Fail (0)** **29 0** **15 0** **33 1** **28 1** **39 1**	10	CO2	BT3
11	Differentiate **Linear SVM** and **Non-Linear SVM** with real-time examples.	10	CO2	BT4

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	**A classification model in Machine Learning produced the following Confusion Matrix results for email spam detection.** **Actual / Predicted Spam Not Spam** **Spam 40 10** **Not Spam 5 45** **Using the above confusion matrix, compute Accuracy, Precision, Recall and F1 Score.**	10	CO2	BT3
13	**A dataset contains the following training samples with two features. Using the K-Nearest Neighbors algorithm, classify the new point P(5,4). Assume K = 3 and use Euclidean distance.** **Data Point Class** **(2,3) A** **(3,4) A** **(4,7) B** **(6,5) B** **(7,3) B**	10	CO2	BT3