

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



PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – May/June 2026

Date: 15- 05-2026

Time: 09:30am – 12:30pm

School: SOIS	Program: 4BCAAIML/DS	
Course Code: CSA2517	Course Name: Machine Learning Algorithms	
Semester: IV	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04	C05	C06
Marks	22	14	22	14	14	14

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.	What is a supervised learning method?	2 Marks	L2	C01
2.	Define training and testing data.	2 Marks	L2	C02
3.	State the difference between classification and regression in Machine Learning.	2 Marks	L2	C03
4.	What is PCA ?	2 Marks	L2	C02
5.	List the types of approaches in hierarchical clustering.		L2	C04
6.	Write a formula for accuracy.	2 Marks	L2	C05
7.	List the ethical issues in Machine Learning.	2 Marks	L2	C05
8.	Mention the typical steps involved in a basic machine learning workflow using Python.	2 Marks	L2	C04
9.	List the ML Python libraries.	2 Marks	L2	C06
10.	What is Scikit-Learn?	2 Marks	L2	C06

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Discuss the different types of machine learning and summarize their characteristics with examples.	10 Marks	L2	C01
	b.	Use machine learning concepts to classify emails as spam or not spam using suitable features.	10 Marks	L3	C03
Or					
12.	a.	Explain the concepts of features and labels and illustrate their role in prediction models.	10 Marks	L2	C01
	b.	Identify how machine learning models learn from data and summarize the overall workflow.	10 Marks	L3	C03

13.	a.	Describe different encoding techniques and summarize their role in handling categorical data with examples.	10 Marks	L2	C02																					
	b.	Evaluate the performance of logistic regression and calculate the metrics Accuracy, Precision, Recall and F1- Score and predict whether a student pass or fail on the given dataset. Given slope = 0.8, intercept= - 0.8 <table><tr><th>Student</th><th>Hours (X)</th><th>Actual (Y)</th></tr><tr><td>1</td><td>1</td><td>0</td></tr><tr><td>2</td><td>2</td><td>0</td></tr><tr><td>3</td><td>3</td><td>0</td></tr><tr><td>4</td><td>4</td><td>1</td></tr><tr><td>5</td><td>5</td><td>1</td></tr><tr><td>6</td><td>6</td><td>1</td></tr></table>	Student	Hours (X)	Actual (Y)	1	1	0	2	2	0	3	3	0	4	4	1	5	5	1	6	6	1	10 Marks	L5	C05
Student	Hours (X)	Actual (Y)																								
1	1	0																								
2	2	0																								
3	3	0																								
4	4	1																								
5	5	1																								
6	6	1																								

Or																							
14.	a.	Explain the concept of cross-validation and summarize how it improves model reliability with examples.	10 Marks	L2	C02																		
	b.	Evaluate the performance of logistic regression and calculate the metrics Accuracy, Precision, Recall and F1- Score and predict whether a patient is diabetic or non-diabetic on the given dataset. Given slope = 0.1, intercept= -10 <table><tr><th>Patient</th><th>Glucose (X)</th><th>Actual (Y)</th></tr><tr><td>1</td><td>85</td><td>0</td></tr><tr><td>2</td><td>90</td><td>0</td></tr><tr><td>3</td><td>95</td><td>1</td></tr><tr><td>4</td><td>100</td><td>1</td></tr><tr><td>5</td><td>105</td><td>1</td></tr></table>	Patient	Glucose (X)	Actual (Y)	1	85	0	2	90	0	3	95	1	4	100	1	5	105	1	10 Marks	L5	C05
Patient	Glucose (X)	Actual (Y)																					
1	85	0																					
2	90	0																					
3	95	1																					
4	100	1																					
5	105	1																					

15.	a.	Explain the types and methods of clustering techniques.	10 Marks	L2	C01																		
	b.	Apply K-Means clustering for the following data points for K = 2 and show all the iterations: Data: 2, 4, 8, 10, 12, 14.	10 Marks	L3	C03																		
Or																							
16.	a.	Describe the working of hierarchical clustering and interpret the significance of a dendrogram in cluster formation with an example.	10 Marks	L2	C01																		
	b.	Implement PCA on the given dataset and identify the new features. <table><tr><th>Student</th><th>Math Marks (X1)</th><th>Physics Marks (X2)</th></tr><tr><td>A</td><td>90</td><td>85</td></tr><tr><td>B</td><td>85</td><td>80</td></tr><tr><td>C</td><td>80</td><td>78</td></tr><tr><td>D</td><td>70</td><td>65</td></tr><tr><td>E</td><td>60</td><td>60</td></tr></table>	Student	Math Marks (X1)	Physics Marks (X2)	A	90	85	B	85	80	C	80	78	D	70	65	E	60	60	10 Marks	L3	C03
Student	Math Marks (X1)	Physics Marks (X2)																					
A	90	85																					
B	85	80																					
C	80	78																					
D	70	65																					
E	60	60																					

17.	a.	Examine the process of training and testing datasets in building a classification model. Explain each step, including dataset preparation, train-test split, model training and performance metrics on the given dataset	10 Marks	L4	C04																		
<table><tr><th>Income (₹)</th><th>Credit Score</th><th>Loan Approved</th></tr><tr><td>20000</td><td>600</td><td>0</td></tr><tr><td>25000</td><td>650</td><td>0</td></tr><tr><td>30000</td><td>700</td><td>1</td></tr><tr><td>35000</td><td>720</td><td>1</td></tr><tr><td>40000</td><td>750</td><td>1</td></tr></table>						Income (₹)	Credit Score	Loan Approved	20000	600	0	25000	650	0	30000	700	1	35000	720	1	40000	750	1
Income (₹)	Credit Score	Loan Approved																					
20000	600	0																					
25000	650	0																					
30000	700	1																					
35000	720	1																					
40000	750	1																					
	b.	Discuss a mini case study on classification using a student performance dataset. Explain the problem statement, approach, and outcome.	10 Marks	L5	C06																		
Or																							
18.	a.	Examine the process of training and testing datasets in building a regression model. Explain each step, including dataset preparation, train-test split, model training, and performance metrics on the given dataset.	10 Marks	L4	C04																		
<table><tr><th>Experience (Years)</th><th>Salary (₹)</th></tr><tr><td>1</td><td>150</td></tr><tr><td>2</td><td>180</td></tr><tr><td>3</td><td>250</td></tr><tr><td>4</td><td>300</td></tr><tr><td>5</td><td>350</td></tr></table>						Experience (Years)	Salary (₹)	1	150	2	180	3	250	4	300	5	350						
Experience (Years)	Salary (₹)																						
1	150																						
2	180																						
3	250																						
4	300																						
5	350																						
	b.	Discuss a mini case study on classification using a diabetic dataset. Explain the problem statement, approach, and outcome.	10 Marks	L5	C06																		

