



# PRESIDENCY UNIVERSITY

BENGALURU

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

## End - Term Examinations - December 2025

Date: 10- 12- 2025

Time: 1.00pm to 04.00pm

<b>School:</b> SOCSE	<b>Program:</b> B.Tech CSD		
<b>Course Code:</b> ADS2004	<b>Course Name:</b> Machine Learning for Intelligent Data Science		
<b>Semester:</b> V	<b>Max Marks:</b> 100	<b>Weightage:</b> 50%	

CO - Levels	C01	C02	C03	C04
Marks	24	24	26	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.	Identify key tools that support data science workflows.	2 Marks	L1	C01
2.	Define correlation and list the types of correlation.	2 Marks	L1	C01
3.	What is the definition of Recall in classification?	2 Marks	L1	C02
4.	State the formula for calculating the F1 Score.	2 Marks	L1	C02
5.	Substitute the given values ( $W = 227$ , $K = 11$ , $P = 0$ , $S = 4$ ) into the formula.	2 Marks	L1	C03
6.	Identify the main reason why CNNs are preferred over ANNs for image recognition tasks.	2 Marks	L1	C03
7.	What is meant by stride = 2 in convolution?	2 Marks	L1	C03
8.	List the key components of the Hadoop framework.	2 Marks	L1	C04
9.	State the differences between RHadoop and Rhipe in the Hadoop ecosystem.	2 Marks	L1	C04
10.	Why might SparkR be preferred for big data analysis?	2 Marks	L1	C04

## Part B

**Answer the Questions.**

**Total Marks 80M**

<b>11.</b>	<b>a.</b>	Explain how the overlaps between different fields or concepts create new insights or opportunities.	<b>20 Marks</b>	<b>L2</b>	<b>CO1</b>
<b>Or</b>					
<b>12.</b>	<b>a.</b>	Summarize the characteristics of effective data analytic thinking.	<b>20 Marks</b>	<b>L2</b>	<b>CO1</b>

<b>13.</b>	<b>a.</b>	<p>A hospital wants to predict patient outcomes:</p> <p>Type 0: Whether a patient will be readmitted within 30 days (classification).</p> <p>Type 1: Predict the number of days a patient will stay in the hospital (regression).</p> <p style="margin-left: 40px;"><b>i)</b> Apply an appropriate ensemble classifier (e.g., Random Forest or AdaBoost) for the Type 0 problem.</p> <p style="margin-left: 40px;"><b>ii)</b> Apply an appropriate ensemble regressor (e.g., Gradient Boosting or Bagged Regression Trees) for the Type 1 problem.</p> <p style="margin-left: 40px;"><b>iii)</b> Evaluate both models and explain how ensemble learning improves prediction accuracy compared to single models.</p>	<b>20 Marks</b>	<b>L3</b>	<b>CO2</b>
------------	-----------	---	-----------------	-----------	------------

**Or**

<b>14.</b>	<b>a.</b>	<p>A company has a dataset of customer data including features like age, income, and purchase history. The goal is to predict whether a customer will respond positively to a new marketing campaign (binary classification).</p> <p style="margin-left: 40px;"><b>i)</b> Apply a Support Vector Machine (SVM) classifier to this dataset.</p> <p style="margin-left: 40px;"><b>ii)</b> Experiment with different kernel functions (linear, polynomial, RBF) and compare their performance.</p> <p style="margin-left: 40px;"><b>iii)</b> Explain which kernel is most suitable for this dataset and why, based on model accuracy and decision boundary.</p>	<b>20 Marks</b>	<b>L3</b>	<b>CO2</b>
------------	-----------	--	-----------------	-----------	------------

<b>15.</b>	<b>a.</b>	<p>Given four different real-world sequence processing tasks — text classification, music generation, sentiment analysis, and machine translation — identify and apply the appropriate RNN architecture type (one-to-one, one-to-many, many-to-one, or many-to-many) for each task. Justify your choice with a brief</p>	<b>20 Marks</b>	<b>L3</b>	<b>CO3</b>
------------	-----------	--	-----------------	-----------	------------

		explanation of how the input and output sequences are structured in each case.			
<b>Or</b>					
<b>16.</b>	<b>a.</b>	Given a sequence prediction problem where long-term dependencies exist but computational efficiency is also important, apply the Gated Recurrent Unit (GRU) architecture to model the sequence. Explain how the update gate and reset gate in GRU help manage information flow compared to a standard RNN, and describe how you would apply GRU to predict future stock prices based on past data.	<b>20 Marks</b>	<b>L3</b>	<b>CO3</b>
<b>Or</b>					
<b>17.</b>	<b>a.</b>	Develop a simple R script that connects to Hadoop for performing data aggregation or transformation tasks.	<b>20 Marks</b>	<b>L3</b>	<b>CO4</b>
<b>Or</b>					
<b>18.</b>	<b>a.</b>	Construct a data visualization using the ggplot2 package in R by creating a graph that represents the relationship between two variables, and explain how each component of the ggplot function contributes to the final output.	<b>20 Marks</b>	<b>L3</b>	<b>CO4</b>