



PRESIDENCY UNIVERSITY

BENGALURU

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

Make Up Examinations – December 2025

Date: 26 – 12- 2025

Time: 1.00pm to 04.00pm

School: SOIS	Program: BCA		
Course Code: CSA1003	Course Name: FUNDAMENTAL OF DATA SCIENCE		
Semester: MK	Max Marks: 100	Weightage: 50%	

CO - Levels	C01	C02	C03	C04	C05
Marks	30	20	20	30	

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.	Define Machine Learning.	2 Marks	L1	C01
2.	What are two major benefits of using data science in business decision-making?	2 Marks	L1	C01
3.	List any four stages in the Data Science Process.	2 Marks	L1	C01
4.	What are the three main uses of data science in real-world applications?	2 Marks	L1	C01
5.	What is the purpose of exploratory data analysis (EDA)?	2 Marks	L1	C01
6.	qualitative and quantitative data with one example each.	2 Marks	L2	C01
7.	What do you mean by frequency table in data description?	2 Marks	L1	C01
8.	Write any two measures used to describe variability in data.	2 Marks	L1	C01
9.	Difference between data engineer and data scientist.	2 Marks	L2	C01
10.	Specify name of any two supervised learning techniques	2 Marks	L1	C01

Part B

Answer the Questions.

Total Marks 80M

11.	a.	<p>For the data set [4, 5, 9, 6, 2, 18, 8, 5, 8, 7, 8, 3, 17], calculate all three.</p> <p>Calculate following:</p> <p style="margin-left: 20px;">I. Mean II. Median III. Mode</p>	20 Marks	L3	CO2
	b.	<p>Calculate standard deviation for the data set [4, 6, 8, 6, 2, 8, 8, 5, 7, 3, 7]. (Each and all steps are to be followed).</p>		L3	CO2

Or

12.	a.	<p>Obtain X and Y. Analyze which is best x y, Variance for the following observation</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 50%;">X</th> <th style="width: 50%;">Y</th> </tr> </thead> <tbody> <tr><td>10</td><td>42</td></tr> <tr><td>16</td><td>10</td></tr> <tr><td>21</td><td>11</td></tr> <tr><td>10</td><td>34</td></tr> <tr><td>24</td><td>24</td></tr> <tr><td>22</td><td>97</td></tr> <tr><td>19</td><td>14</td></tr> </tbody> </table>	X	Y	10	42	16	10	21	11	10	34	24	24	22	97	19	14	20 Marks	L3	CO2
X	Y																				
10	42																				
16	10																				
21	11																				
10	34																				
24	24																				
22	97																				
19	14																				

13.	a.	<p>You are given a document with the following lines:</p> <p>Line 1: Hadoop MapReduce is powerful Line 2: Hadoop is scalable Line 3: MapReduce simplifies big data processing</p> <p>Write the Map and Reduce output to compute the word frequency count using the MapReduce paradigm.</p>	20 Marks	L4	CO3
	b.	<p>Illustrate the concept of regression towards the mean using a real-life or hypothetical example. Why is this phenomenon important in understanding prediction errors?</p>		L2	CO3

Or

14.	a.	<p>Given a dataset: Use the least squares method to calculate the best-fit linear regression line ($Y = a + bX$) and predict the score if a student studies for 6 hours. Use the least squares method to calculate the best-fit linear regression line ($Y = a + bX$) and</p>	20 Marks	L3	CO3
------------	-----------	---	-----------------	-----------	------------

		predict the score if a student studies for 6 hours.															
		<table border="1"> <thead> <tr> <th>X (Hours Studied)</th> <th>Y (Score)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>40</td> </tr> <tr> <td>2</td> <td>50</td> </tr> <tr> <td>3</td> <td>60</td> </tr> <tr> <td>4</td> <td>65</td> </tr> <tr> <td>5</td> <td>70</td> </tr> </tbody> </table>	X (Hours Studied)	Y (Score)	1	40	2	50	3	60	4	65	5	70			
X (Hours Studied)	Y (Score)																
1	40																
2	50																
3	60																
4	65																
5	70																

15.	a.	<p>You are given two documents:</p> <ul style="list-style-type: none"> Doc1: "machine learning is fun" Doc2: "learning machine algorithms is essential" <p>Treat each document as a vector of word frequencies. Calculate the Cosine Similarity between the two documents to determine how similar they are.</p>	20 Marks	L4	C04
	b.	Explain EDA and its techniques.		L1	C01

Or

16.	a.	<p>Obtain X and Y. Analyze which is best x y, Standard deviation for the following observation:</p> <table border="1"> <thead> <tr> <th>X</th> <th>Y</th> </tr> </thead> <tbody> <tr> <td>15</td> <td>72</td> </tr> <tr> <td>26</td> <td>41</td> </tr> <tr> <td>21</td> <td>10</td> </tr> </tbody> </table>	X	Y	15	72	26	41	21	10	20 Marks	L3	C04
X	Y												
15	72												
26	41												
21	10												

17.	a.	<p>Compare the following:</p> <ol style="list-style-type: none"> Regression and Classification Reinforcement Learning and online learning. 	20 Marks	L2	C04
------------	-----------	--	-----------------	-----------	------------

Or

18.	a.	<p>Demonstrate visualization through following plots by assuming suitable dataset:</p> <ol style="list-style-type: none"> scatter plot pair plot count plot heatmap 	20 Marks	L4	C03
------------	-----------	---	-----------------	-----------	------------