Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – MAY 2025

Date: 20-05-2025 **Time:** 09:30 am – 12:30 pm

School: SOCSE	**Program:** B.Tech-CSD	
Course Code: CSE3038	**Course Name:** Applied Data Science	
Semester: VI	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4	CO5
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.	What is the difference between qualitative and quantitative data?	2 Marks	L1	CO1
2.	Mention any two Python libraries commonly used in data science and their purpose.	2 Marks	L1	CO1
3.	What is the role of feature scaling in data preprocessing?	2 Marks	L2	CO2
4.	Explain one-hot encoding with a use case.	2 Marks	L2	CO2
5.	What are support vectors in SVM?	2 Marks	L1	CO3
6.	Define precision and recall.	2 Marks	L2	CO3
7.	List any two applications of supervised learning.	2 Marks	L1	CO3
8.	What is a medoid in K-Medoids clustering?	2 Marks	L1	CO4
9.	List any two clustering algorithms other than K-Means.	2 Marks	L1	CO4
10.	Why is clustering considered an unsupervised technique?	2 Marks	L2	CO4

Part B

Answer the Questions. **Total Marks 80M**

11.	a.	Explain the types of Python operators with examples.	10 Marks	L2	CO1
	b.	Write a Python program to accept a sentence from the user and count the number of vowels in it.	10 Marks	L2	CO1
		Or			
12.	a.	Describe different data types in Python with suitable examples.	10 Marks	L2	CO1
	b.	Write a program to perform basic NumPy operations on a one-dimensional array (addition, multiplication, mean).	10 Marks	L2	CO1

13.	a.	What are the statistical descriptors used to describe a dataset?	10 Marks	L1	CO2
	b.	Create a pair plot and scatter plot using matplotlib for a given dataset & describe it.	10 Marks	L2	CO2
		Or			
14.	a.	Describe the different types of colormaps in matplotlib and their significance.	10 Marks	L2	CO2
	b.	Write a program to read an XML dataset in Python and print selected fields.	10 Marks	L2	CO2

15.	a.	Explain the differences between decision tree and naive Bayes classifiers.	10 Marks	L2	CO3
	b.	Apply SVM using Scikit-learn with a non-linear kernel on a UCI dataset.	10 Marks	L3	CO3
		Or			
16.	a.	Demonstrate the overfitting concept and how to handle it in supervised learning.	10 Marks	L3	CO3
	b.	Restructure a Naive Bayes classifier and evaluate its performance using confusion matrix.	10 Marks	L3	CO3

17.	a.	Explain the steps involved in the DBSCAN algorithm.	10 Marks	L2	CO4
	b.	Apply K-Medoids clustering on a small dataset and calculate cost.	10 Marks	L3	CO4
		Or			
18.	a.	Describe the Elbow Method for selecting the number of clusters.	10 Marks	L2	CO4
	b.	Apply hierarchical clustering (agglomerative) on a dataset and plot the dendrogram.	10 Marks	L3	CO4