

STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI – 600 086
(For candidates admitted from the academic year 2024 – 2025 and thereafter)

M.Sc. DEGREE EXAMINATION – APRIL 2026
DATA SCIENCE
SECOND SEMESTER

COURSE : CORE

PAPER : MACHINE LEARNING – THEORY AND PRACTICE

SUBJECT CODE : 24DS/PC/MA24

TIME : 1 ½ HOURS

MAX. MARKS: 50

SECTION - A			
Answer all the questions		(6 x 5 = 30)	
Q.No		CO	CL
1.	a) List the main steps that are required to perform Linear Discriminant Analysis. (OR) b) Enumerate on dealing with missing data.	CO1	K1
2.	a) Explain the procedure involved in boosting algorithm. (OR) b) Discuss the steps involved in the k-means clustering algorithm.	CO2	K2
3.	a) Demonstrate why using a numpy array is more efficient than a python list. Given a list of numbers, convert the list into a numpy array and computer any two statistical measures using numpy methods. (OR) b) Demonstrate how the pandas methods can be used to gain data insights from a CSV file.	CO3	K3
4.	a) Illustrate streamlining workflows with Pipeline and explain. (OR) b) Explain the steps necessary to implement the kernel principal component analysis.	CO3	K3
5.	a) Analyse the three different types of Machine Learning. (OR) b) Explain bias-variance tradeoff and analyse how regularization can help in controlling the model complexity to achieve a good bias–variance tradeoff in machine learning models.	CO4	K4
6.	a) Examine and explain the working of bootstrap aggregating of a bagging classifier with an example. (OR) b) Examine how the elbow method and silhouette plots help evaluate the quality of clustering results.	CO4	K4

SECTION - B			
Answer all the questions		(2 x 10 = 20)	
7.	a) Illustrate and explain in detail how holdout cross-validation and k-fold cross-validation can help estimate the model's performance on unseen data. (OR) b) Explain how a trained machine learning model can be embedded into a web application. Briefly describe the steps involved in deploying the model and enabling the application to learn from data in real time.	CO3	K3
8.	a) Illustrate the typical workflow used in Machine Learning for predictive modeling with a neat diagram. Analyse the workflow and explain the role of each stage involved. (OR) b) Analyse how logistic regression can be used as a probabilistic model for binary classification. Explain the role of the sigmoid function and sigmoidal curve and describe how they help predict the probability of a data sample belonging to a particular class.	CO4	K4
