

Sri Sivasubramaniya Nadar College of Engineering, Chennai
(Autonomous Institution under Anna University)

Degree & Branch	5 years Integrated M.Tech CSE	Semester	V
Subject Code & Name	ICS1512 – Machine Learning Algorithms Laboratory		
Academic Year	2025–2026 (Odd Semester)	Batch	2023–2028
Name	Pravin G	Reg No	3122237001041

Experiment #1: Working with Python Packages – Numpy, Scipy, Scikit-learn, Matplotlib

Aim:

To explore python libraries to perform EDA, data preprocessing, data splitting, and performance evaluation on real-life datasets.

Libraries used:

- Numpy
- Pandas
- Scipy
- Scikit-Learn
- Matplotlib.pyplot

Description of the objective performed

- **Data Preprocessing:** Loaded dataset using `sklearn.datasets.load_iris()` and converted it into a Pandas DataFrame.
- **Exploratory Data Analysis (EDA):**
 - Calculated summary statistics using `pandas.describe()`
 - Visualized distributions and relationships using:
 - * `plt.hist()` for histograms
 - * `plt.scatter()` for 2D scatter plots
 - * `sns.heatmap()` for feature correlation matrix
- **Classification using k-Nearest Neighbors (KNN):**
 - Distance Metric: Euclidean distance
 - For a new point x , find k nearest neighbors and assign the most frequent class.
 - Accuracy measured using `accuracy.score()` from Scikit-learn.

Results and Discussions:

Dataset	Type of ML Task	Suitable ML Algorithm
Iris Dataset	Classification	k-Nearest Neighbors (KNN)
Loan Amount Prediction	Regression	Linear Regression
Predicting Diabetes	Classification	Logistic Regression
Classification of Email Spam	Classification	Naive Bayes, Support Vector Machine (SVM)
Handwritten Character Recognition / MNIST	Classification	Convolutional Neural Networks (CNN)

Table 1: Types of ML tasks, and suitable algorithms for various datasets

Learning Outcomes:

- Able to differentiate between classification and regression problems and choose suitable machine learning algorithms based on the nature of the task and dataset characteristics.
- Able to utilize Numpy and Pandas for data handling, and preprocessing techniques
- Able to apply data visualization techniques using Matplotlib and Seaborn to perform exploratory data analysis (EDA), enabling the identification of patterns, trends, and correlations within datasets.