

Sri Sivasubramaniya Nadar College of Engineering, Chennai
(An autonomous Institution affiliated to Anna University)

Degree & Branch	B.E Computer Science & Engineering	Semester	VI
Subject Code & Name	UCS2612 & Machine Learning Algorithms Laboratory		
Academic year	2025-2026 Even	Batch:2023-2027	Due date:

Experiment 1: Working with Python packages-Numpy, Scipy, Scikit-Learn, Matplotlib

1. Explore the various functions and methods available in the following Python libraries: **Numpy, Pandas, Scipy, Scikit-learn, Matplotlib**. Understand the key operations such as array manipulations, data preprocessing, mathematical computing, machine learning workflows, and data visualization [CO1, K2]. **Helpful Resources:**
 - Numpy: Official Tutorials, W3Schools Guide
 - Pandas: Official 10-Min Guide, W3Schools Guide
 - Scipy: Official Special Functions, W3Schools Graphs
 - Scikit-learn: Official Site, Tutorials
 - Matplotlib: Official Tutorials, W3Schools Intro
2. Explore public repositories such as the UCI Machine Learning Repository (UCI Repository) and Kaggle Datasets. Download the following datasets and identify the appropriate machine learning model to be used (e.g., Supervised, Unsupervised, Semi-supervised, Regression, Classification) [CO1, K3].
 - i.) Loan amount prediction
 - ii.) Handwritten character recognition
 - iii.) Classification of Email spam and MNIST data
 - iv.) Predicting Diabetes
 - v.) Iris Dataset
3. Identify the type of ML task associated with each dataset. Explore the steps involved in the machine learning workflow [CO1, K2].
 - i.) **Loading the dataset**
 - ii.) **Exploratory Data Analysis and Visualization:** Analyze data using summary statistics and visual tools such as histograms, bar charts, scatter plots, heatmaps, box plots, etc.
 - iii.) **Data Preprocessing:** Handle missing values, drop irrelevant features, encode categorical variables, normalize/standardize data.

- iv.) **Feature Selection:** Use techniques like SelectKBest, Chi-square test, and ANOVA to select relevant features.
- v.) **Data Splitting:** Divide the dataset into training, testing, and validation sets.
- vi.) **Performance Evaluation:** Study appropriate metrics for both supervised and unsupervised learning models and use visualizations to interpret results.

Identify the type of task performed for above datasets and tabulated the details in the below table.

Dataset	Type of ML task	Feature Selection Technique	Suitable ML Algorithm
Iris Dataset			
Loan Amount Prediction			
Predicting Diabetes			
Classification of Email Spam			
Handwritten Character Recognition / MNIST			

Note:

A supplementary document on *Exploratory Data Analysis (EDA)* has been uploaded to the LMS portal to support better understanding of the concepts and visualizations discussed in this experiment.

Optional Task:

Explore and experiment with advanced deep learning libraries such as Theano, TensorFlow, Keras, and PyTorch.

Submission Guidelines

- **Observation Note:** Include the aim, a summary table of tasks, and key learning outcomes.
- **In LMS (Single PDF Upload):**
 - Clearly state the aim of the experiment.
 - Include well-commented Python code.
 - Capture and include output screenshots.
 - Summarize findings in an inference table.
 - Include a reflection on learning outcomes.