



CHITTAGONG UNIVERSITY OF ENGINEERING & TECHNOLOGY

*Department of*

COMPUTER SCIENCE & ENGINEERING

**Course Title:** Machine Learning (Sessional)

**Course Code:** CSE-464

**Report No:** 01

**Report Title:** Basic Implementation of NumPy, Pandas & Matplotlib

**Name:** Tahlil Abrar

**ID:** 1804056

**Section:** A

**Submitted to:**

1. Md. Rashadur Rahman

Lecturer, Dept. Of CSE, CUET

2. Hasan Murad

Lecturer, Dept. Of CSE, CUET

Remarks

## **Objective:**

1. To familiarize with the basic concepts of NumPy, Pandas, and Matplotlib
2. To perform the basic implementation of these libraries

## **Introduction:**

NumPy, Pandas, and Matplotlib are essential libraries in the field of Machine Learning (ML) for data manipulation, analysis, and visualization. Together, NumPy, Pandas, and Matplotlib form a powerful ecosystem for data manipulation, analysis, and visualization, enabling ML practitioners to handle datasets, pre-process data, explore features, and create informative visualizations for model development and evaluation. These libraries are widely used by data scientists, researchers, and analysts because they provide powerful functionality and are easy to use. NumPy is a fundamental library for scientific computing in Python. It provides powerful tools for working with large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to perform operations efficiently. NumPy is the foundation for many other libraries in the Python data ecosystem, including Pandas and Matplotlib. Pandas is a popular data manipulation library in Python. It offers data structures and functions for efficiently handling and analyzing structured data, primarily in tabular form. The two primary data structures in Pandas are Series (1D labeled array) and DataFrame (2D labeled array), which provide powerful tools for data manipulation, cleaning, and exploration. Matplotlib is a versatile plotting library in Python that enables the creation of various types of visualizations, ranging from simple line plots to complex figures. It provides a wide range of customization options and supports different output formats.

## **NumPy:**

NumPy is a fundamental library for scientific computing in Python. It provides powerful tools for working with large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to perform operations efficiently. NumPy is the foundation for many other libraries in the Python data ecosystem, including Pandas and Matplotlib.

In ML, NumPy is extensively used for tasks such as:

- ❖ Handling and manipulating numerical data efficiently.
- ❖ Performing mathematical operations on arrays, such as element-wise operations, linear algebra, and statistical calculations.
- ❖ Creating and manipulating feature matrices and target arrays for ML models.

- ❖ Supporting vectorized computations, improving computational efficiency.

## **Pandas:**

Pandas is a popular data manipulation library in Python. It offers data structures and functions for efficiently handling and analyzing structured data, primarily in tabular form. The two primary data structures in Pandas are Series (1D labeled array) and DataFrame (2D labeled array), which provide powerful tools for data manipulation, cleaning, and exploration.

In ML, Pandas is commonly used for:

- ❖ Loading, preprocessing, and cleaning data from various sources, such as CSV files, databases, or APIs.
- ❖ Handling missing values, outliers, and data inconsistencies.
- ❖ Performing data exploration and descriptive statistics.
- ❖ Merging, joining and reshaping datasets.
- ❖ Preparing data for ML models by transforming, encoding, or scaling features.

## **Matplotlib:**

Matplotlib is a versatile plotting library in Python that enables the creation of various types of visualizations, ranging from simple line plots to complex figures. It provides a wide range of customization options and supports different output formats.

In ML, Matplotlib is useful for:

- ❖ Visualizing data distributions, trends, and relationships.
- ❖ Creating scatter plots, bar plots, histograms, box plots, and other types of visualizations.
- ❖ Evaluating model performance through plots like ROC curves or confusion matrices.
- ❖ Analyzing model outputs, such as feature importance or decision boundaries.
- ❖ Communicating and presenting results effectively.

**Discussion:**

In this report, we have observed the basic implementation of three useful Python libraries and learned how they are used to manipulate or visualize data. By combining the power of NumPy, Pandas, and Matplotlib, users can efficiently process and analyze data while creating informative and visually appealing visualizations. These libraries form the backbone of the Python data system and are essential for any data scientist or analyst.