

Lab: Python Numpy for Machine Learning

- [NumPy](#) is a Python package which means 'Numerical Python'.
 - Numerical data is data in the form of numbers.
- It provides us with a powerful object known as an [Array](#). Numpy array is a powerful N-dimensional array object which is in the form of rows and columns.
- With Arrays, we can perform mathematical operations on multiple values in the Arrays at the same time, and perform operations between different Arrays, similar to matrix operations.
- **Source Code**

```
pip install numpy // PIP is a package manager for Python packages, or modules
```

```
import numpy as np
student_information = [['Aman', 95.5, "Male"], ['Sunny', 65.7, "Female"],
                      ['Monty', 85.1, "Male"], ['toni', 75.4, "Male"]]

# Creating a 3-D numpy array using np.array()
student_array=np.array(student_information)

# Printing the Numpy array
print(student_array)
```

- **Output**
[['Aman' '95.5' 'Male']
['Sunny' '65.7' 'Female']
['Monty' '85.1' 'Male']
['toni' '75.4' 'Male']]

Run, Check, Understand:

<https://colab.research.google.com/drive/1Nsa5rVHD6nG9aEdTHsZkYvqUoZ3hmfYq#scrollTo=9vyY8jwq82uE>

Problem 1 (Member 1): Weather Data Analysis (2D Array)

Dataset Type: A 2D array representing weather data, where rows are days and columns are [temperature, humidity, wind speed].

Task:

1. Create a random dataset with 30 days of weather data (shape: 30x3) using `np.random.rand()`. Scale the values:
 - Temperature: 15–35°C
 - Humidity: 30–80%
 - Wind Speed: 0–20 m/s
2. Calculate:
 - Average temperature, humidity, and wind speed over the month.
 - Maximum wind speed and the day it occurred.
 - Number of days with temperatures above 30°C.
3. Extract a subset of data (days 10 to 20) and find the variance for each column.

Problem 2 (Member 2): Product Sales Analysis (Categorical Data in Integer Format)

Dataset Type: A 1D array representing the product IDs of 1000 sales transactions.

Task:

1. Generate a random dataset of integers (product IDs ranging from 1 to 10) using `np.random.randint()`.
2. Calculate:
 - The frequency of each product sold using `np.unique()`.
 - The most sold product and the number of times it was sold.
 - The least sold product.
3. Create a new array where:
 - If the product ID is less than 5, assign it to category 0.
 - If the product ID is 5 or greater, assign it to category 1 using `np.where()`.

Problem 3 (Member 3): Time Series Analysis (Stock Prices)

Dataset Type: A 1D array representing daily closing prices of a stock over 256 days.

Task:

1. Generate a random time series of stock prices (256 days) with values ranging between 100 and 500 using `np.random.uniform()`.
2. Perform the following operations:
 - Normalize the stock prices to a range of 0 to 1.
 - Calculate the mean and standard deviation of the stock prices.
 - Identify and count the number of days where the normalized price is greater than 0.8.
 - Extract a subset of prices for the first 128 days.
 - Apply a transformation where each normalized price is squared.