

Week1 - Introduction to Python and NumPy

DS3010 - Introduction to Machine Learning

OVERVIEW & PURPOSE

This lab is about learning to write programs in Python and familiarizing yourself with NumPy.

Task for the Lab

1. Python Basics

- Write a python program to find the area of a rectangle. Take length and breadth as input from the user.
- Write a python program to find whether a given number is prime or not.
- Write a python program to find distance between two points $(x1, y1)$ and $(x2, y2)$.
- Write a program to print the given pattern using a for loop.

```
* * * * *
* * * *
* * *
* *
*
* *
* * *
* * * *
* * * * *
```

2. Python data structures

- a. Create an empty dictionary named 'student records'
- b. Add 3 student records to the dictionary. Each record should include the student's name as the key and a list of marks in Physics, Chemistry, and Mathematics as the value. For example,

`'John': [85, 92, 78]`

`'David': [76, 88, 90]`

`'George': [52, 68, 84]`
- c. Print the student records in a readable format, showing each student's name and their respective marks in Physics, Chemistry, and Mathematics.
- d. Calculate and print the average marks for each student.
- e. Find and print the student with the highest total marks.
- f. Use a dictionary comprehension to create a new dictionary named failed students that includes students who have failed in at least one subject. A student is considered to have failed if their mark in any subject is below 60.

3. NumPy

You are given the task of analyzing the performance of 100 students in a Machine Learning (ML) and Data Engineering (DE) course. Follow the steps below to perform the analysis using NumPy arrays.

- a. Create 1d NumPy arrays named 'Id', 'ML_score', 'DE_score'. 'Id' array should contain unique sequential values, 'ML_score' and 'DE_score' should contain random integer values ranging from 1 to 100.
- b. Make a single 2d array named 'student_data' from the above 1d arrays with 'Id', 'ML_score' and 'DE_scores' as columns.
- c. Create a new array named avg_score and assign the average values of ML and DL scores of each student. Add this array as a new column into the student_data. Print the 'student_data' array values

of top 5 performers based on avg-score.

- d. add a new column in the 'student_data' which contains grades A, B, C, D, E based on avg_scores ≥ 90 , ≥ 80 , ≥ 70 , ≥ 60 , < 60 .

4. Matplotlib and NumPy

- a. Create the following arrays, 'sales_data': A 4x12 matrix detailing monthly sales quantities of 4 products across 12 months. 'price_per_unit': A 1D array indicating the unit price for each product.
- b. Calculate total revenue for each month by performing matrix operation of above arrays.
- c. Create the following visualizations using Matplotlib: A bar plot comparing the total revenue for each month, and a line plot showing the revenue trend over the 12 months for one specific product (choose any product).