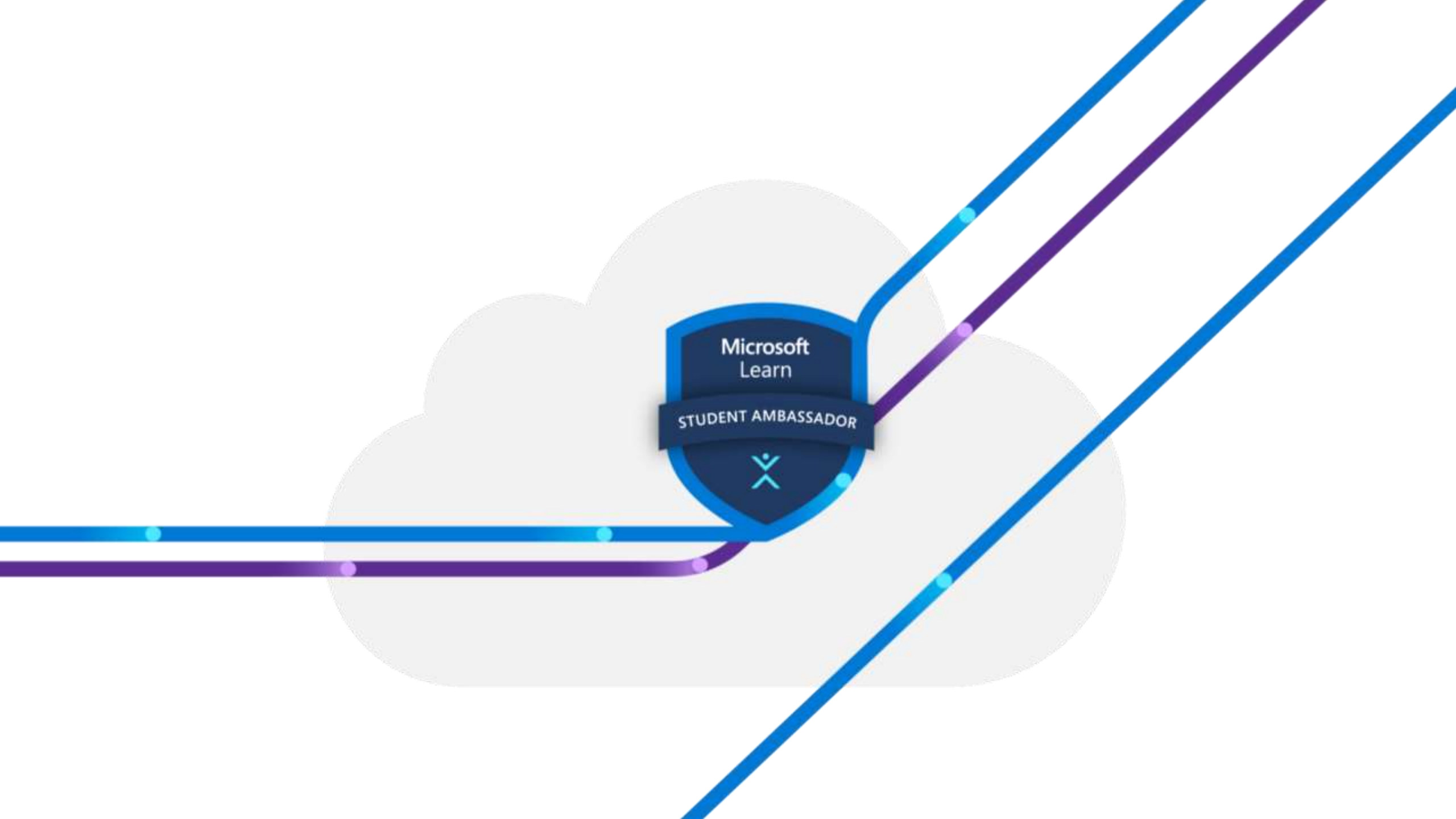

Microsoft
Learn
STUDENT AMBASSADOR

Microsoft Learn
Student Ambassadors

Machine Learning 101 :Beginner's Guide

Mohammed Raza Syed
(Beta Microsoft Learn Student Ambassador)

Topics to be covered :

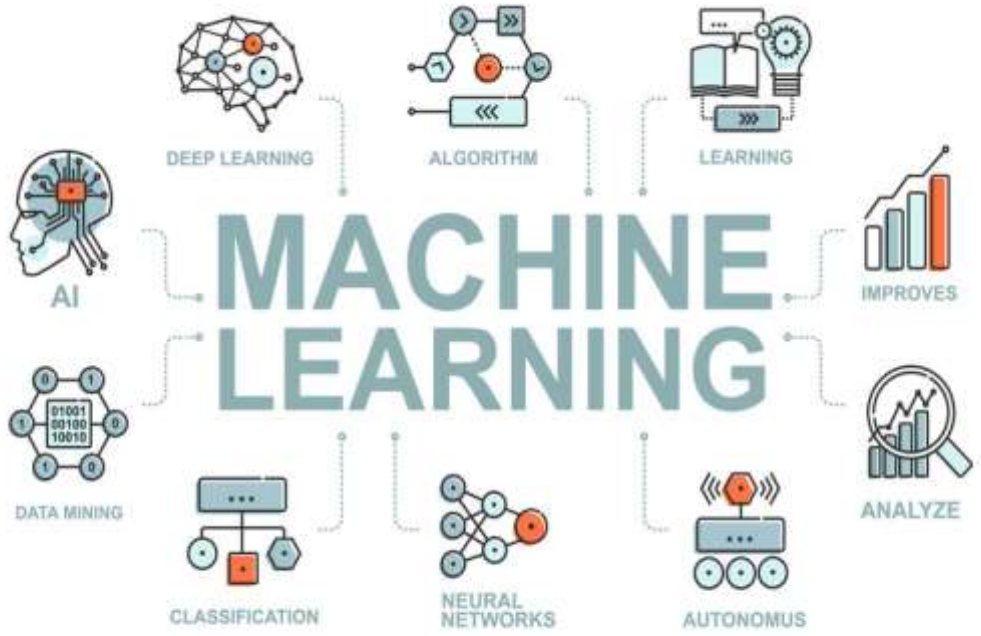
- ❑Introduction to Machine Learning
- ❑Application and Real-World Use Cases of ML
- ❑Introduction to Jupyter Notebooks
- ❑Data Manipulation using Pandas
- ❑Numerical Operation using Numpy
- ❑Data Visualization using Matplotlib
- ❑Quiz
- ❑QnA and Discussions

Introduction to Machine Learning

What is Machine Learning ?

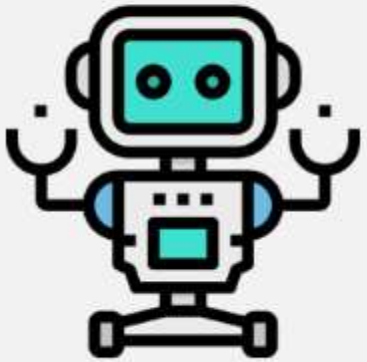
A branch of artificial intelligence that enables systems to learn from data and make predictions or decisions without being explicitly programmed.

Data-driven approach: ML algorithms learn patterns and relationships from data to make predictions or take actions.

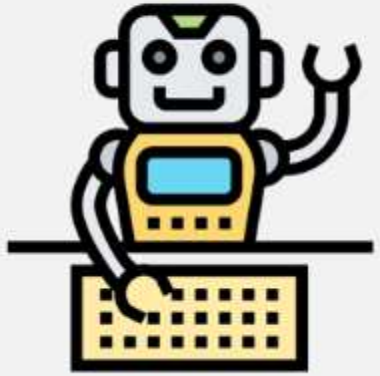


Types of Machine Learning

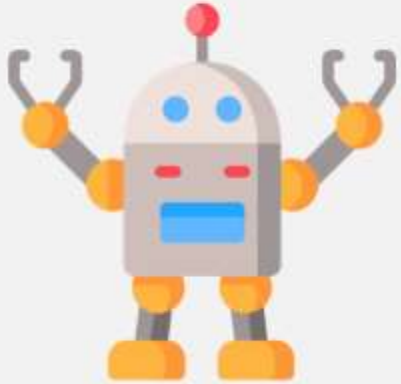
Supervised Learning



Unsupervised Learning



Reinforcement Learning



Machine Learning Algorithms

- **Regression**: Predicting numeric values based on input data.
- **Classification**: Categorizing data into distinct classes or categories.
- **Clustering**: Grouping similar data points together without predefined categories.
- **Dimensionality Reduction**: Reducing the number of input features while retaining important information.
- **Neural Networks**: Creating models inspired by the human brain to learn and make predictions.

Application and Real-World Use cases of ML

Real-World Used Cases of ML

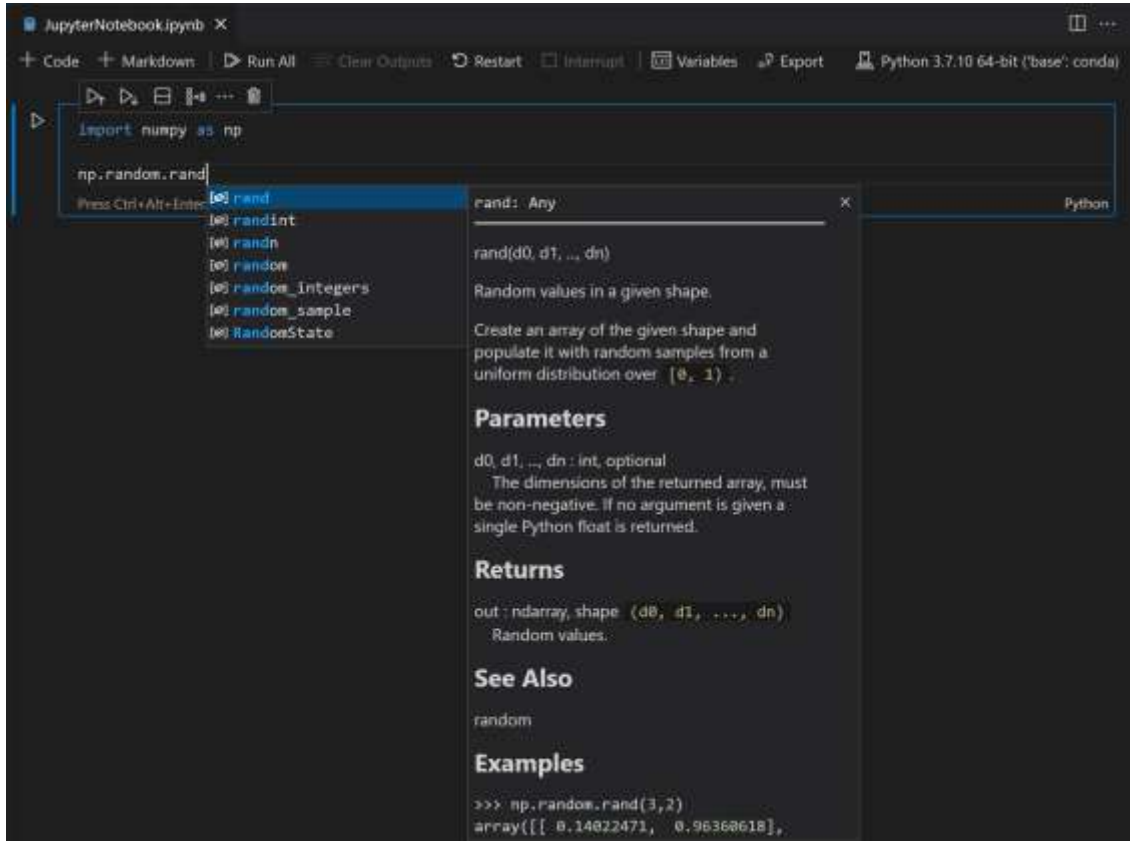
- Healthcare
- Finance
- E-commerce and Marketing
- Autonomous Vehicles
- Natural Language Processing
- Cybersecurity
- and many more…..

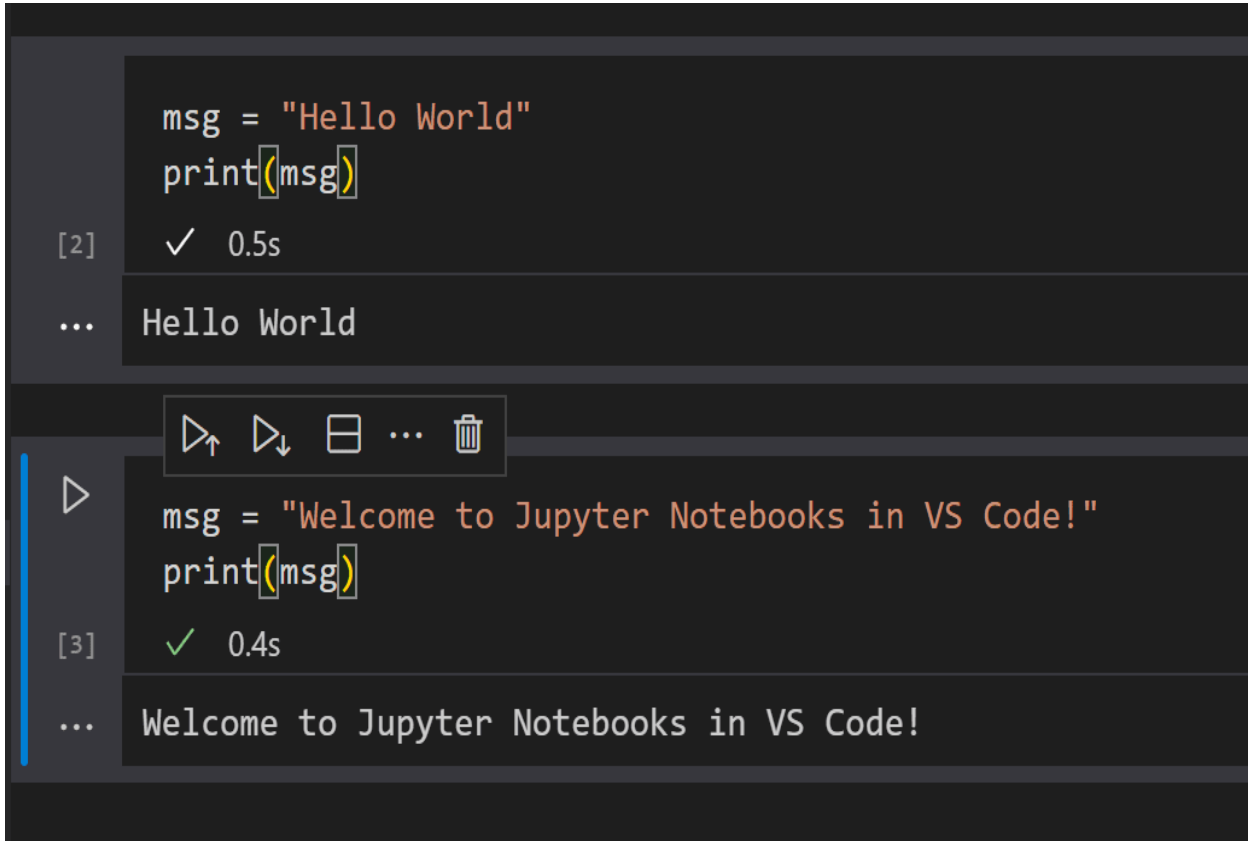
Introduction to Jupyter Notebooks

What are Jupyter Notebooks

- Open-source web app for interactive computing, data analysis, and collaboration
- Create/share documents with live code, visualizations, and text
- Supports Python, R, Julia, and more programming languages
- Ideal for data scientists, researchers, and educators
- Execute code in cells for organization and experimentation
- Combine code, text, and multimedia for data exploration

Working with Jupyter Notebooks in VS Code





Data Manipulation Using Pandas

What is Pandas ?

- Pandas is an open-source data manipulation and analysis library for Python.
- It provides easy-to-use data structures like DataFrames and Series for handling structured data.
- Pandas simplifies tasks such as data cleaning, filtering, transformation, and aggregation.
- It supports various data formats and integrates well with other libraries like NumPy and Matplotlib.
- Pandas offers powerful indexing, slicing, and statistical functions for efficient data analysis and exploration.

Important Functions in Pandas

- read_csv(): Reads a CSV file and returns a DataFrame.
- head(): Returns the first n rows of a DataFrame (default: 5).
- info(): Provides a summary of the DataFrame, including the data types and memory usage.
- describe(): Generates descriptive statistics of numerical columns in the DataFrame.
- value_counts(): Returns the count of unique values in a column.
- dropna(): Drops rows with missing values from a DataFrame.
- isnull(): Identifies and flags null values in a DataFrame.
- groupby(): Groups data based on one or more columns for aggregation or analysis.

Numerical Operation Using Numpy

What is Numpy ?

- NumPy is a fundamental numerical computing library for Python.
- It provides efficient data structures called ndarrays for handling large arrays and matrices.
- NumPy offers a wide range of mathematical functions and operations for array manipulation.
- It enables fast and vectorized computations, improving performance compared to standard Python lists.
- NumPy serves as a foundation for many other libraries in the scientific Python ecosystem, such as Pandas and Matplotlib.

Important Functions in Numpy

- np.array(): Creates a NumPy array from a Python list or tuple.
- np.zeros(): Creates an array filled with zeros.
- np.ones(): Creates an array filled with ones.
- np.random.rand(): Generates an array of random numbers from a uniform distribution.
- np.random.randn(): Generates an array of random numbers from a standard normal distribution.
- np.arange(): Creates an array with regularly spaced values.
- np.reshape(): Changes the shape of an array without modifying its data.
- np.dot(): Performs matrix multiplication between arrays.
- np.sum(): Calculates the sum of array elements along a specified axis.
- np.mean(): Computes the mean value of array elements along a specified axis.

Data Visualization Using Matplotlib

What is Matplotlib ?

- Matplotlib is a popular data visualization library for Python.
- It provides a comprehensive set of tools for creating various types of plots, charts, and graphs.
- Matplotlib offers a user-friendly interface for customizing visualizations with labels, colors, and styles.
- It supports a wide range of plot types, including line plots, bar charts, scatter plots, histograms, and more.
- Matplotlib is highly customizable and allows for interactive visualizations, making it suitable for both exploratory data analysis and presentation purposes.

Important Functions in Matplotlib

- plt.plot(): Plots a line or marker-based graph.
- plt.scatter(): Creates a scatter plot of points.
- plt.bar(): Generates vertical bar plots.
- plt.hist(): Creates histograms to visualize the distribution of data.
- plt.xlabel(): Sets the label for the x-axis.
- plt.ylabel(): Sets the label for the y-axis.
- plt.title(): Sets the title of the plot.
- plt.legend(): Adds a legend to the plot.
- plt.xlim(): Sets the x-axis limits.
- plt.ylim(): Sets the y-axis limits.

Quiz Time !!!!

QnA and Discussions