**ASSIGNMENT-2**

**Scikit-Learn**

**Basic Concepts and Functionalities**

1. **Introduction to Scikit-Learn**:
   * Scikit-Learn is a popular machine learning library in Python that provides simple and efficient tools for data analysis and modeling.
2. **Main Features**:
   * **Consistent Interface**: Provides a uniform interface for various machine learning algorithms (e.g., classification, regression, clustering).
   * **Integration with Numpy and Pandas**: Works seamlessly with Numpy arrays and Pandas DataFrames.
   * **Preprocessing**: Includes tools for data preprocessing such as scaling, normalization, and handling missing values.
   * **Model Selection and Evaluation**: Offers utilities for model selection (cross-validation, grid search) and evaluation (metrics like accuracy, precision, recall).
3. **Basic Functionalities**:
   * **Supervised Learning**: Supports supervised learning algorithms like Linear Regression, Logistic Regression, Support Vector Machines, etc.
   * **Unsupervised Learning**: Provides algorithms for clustering (KMeans, DBSCAN) and dimensionality reduction (PCA).
   * **Pipeline**: Enables constructing pipelines of transformers and estimators, facilitating model building and evaluation.

**Matplotlib vs. Seaborn**

**Comparison for Data Visualization**

**Matplotlib**:

* **Strengths**:
  + Provides extensive customization options for creating static, publication-quality plots.
  + Supports a wide range of plot types and styles, giving full control over every aspect of the plot.
  + Integrates well with other libraries like Pandas and Numpy.
* **Weaknesses**:
  + Requires more code for complex visualizations due to its low-level approach.
  + Default styles may not always be aesthetically pleasing without customization.

**Seaborn**:

* **Strengths**:
  + Simplifies complex plotting tasks with high-level functions, making it easier to create attractive statistical visualizations.
  + Offers built-in themes and color palettes that enhance the visual appeal of plots.
  + Integrates well with Pandas DataFrames and supports plotting categorical data efficiently.
* **Weaknesses**:
  + Limited customization compared to Matplotlib for specific plot details.
  + May not be suitable for highly specialized plots that require fine-tuned control.

**When to Use Each Library**

* **Matplotlib**:
  + Use Matplotlib when you need precise control over every aspect of your plots or when creating complex, customized visualizations.
  + Ideal for creating static, publication-quality plots for presentations and reports.
* **Seaborn**:
  + Choose Seaborn for exploratory data analysis and when you need to quickly generate informative statistical visualizations.
  + Well-suited for tasks like visualizing distributions, correlations, and categorical data.