

Week 1: Introduction to Machine Learning

Task 1.1: Exploring Machine Learning with Python

Objective: Get acquainted with Python tools for data science and explore a basic dataset to understand key statistical concepts.

Dataset: Iris Dataset from Kaggle. This dataset includes 150 samples of iris flowers from three different species along with four features: sepal length, sepal width, petal length, and petal width.

- **Link to dataset:** Iris Dataset on Kaggle : <https://www.kaggle.com/datasets/uciml/iris>

Activities:

1. **Set up the environment:**
 - Ensure Python and Jupyter Notebooks are installed.
 - Install necessary libraries: pandas, matplotlib, seaborn.
 - Load the dataset using pandas.
2. **Data Exploration:**
 - Display the first few rows of the dataset to understand its structure.
 - Use pandas functions like `describe()`, `info()`, and `value_counts()` to gather key statistics and information about the dataset.
3. **Visualizations:**
 - Create histograms for each numerical feature to understand distributions.
 - Use scatter plots to explore potential relationships between features.
 - Employ seaborn's pairplot to visualize the dataset with hue based on the species.

Expected Output:

- A Jupyter notebook that includes:
 - Data loading and exploration steps.
 - Descriptive statistics and interpretations.
 - Visualizations with clear labels and legends.

Documentation:

- Follow the provided documentation template to outline the purpose, process, findings, and any insights from the exploratory data analysis.

General Guidelines for Tasks:

- **Comment your code:** Ensure your code in the Jupyter notebook is well-commented to explain why each step is performed.
- **Consistent Formatting:** Use clear headings and subheadings in your Jupyter notebooks and documentation.

- **Testing and Validation:** After each major step, use simple tests or checks to ensure the transformations are performed as expected.