Exploring and Visualizing the Iris Dataset

Iris Dataset Project Report

1. Objective

The aim of this project is to explore and visualize the Iris dataset. This helps us understand patterns in the data, feature relationships, and spot any outliers. It builds a strong foundation before applying any machine learning models.

2. Dataset Used

Dataset: Iris Flower Dataset

Format: CSV file manually downloaded from Kaggle

Attributes: Sepal length, Sepal width, Petal length, Petal width, Species

3. Libraries Used

- pandas: for data loading and manipulation

- matplotlib: for plotting graphs

- seaborn: for statistical data visualization

4. Steps Performed

1. Loaded the dataset using pandas.

2. Printed dataset shape, column names, and first few rows using .head().

3. Used .info() and .describe() to inspect data and understand summary statistics.

4. Created scatter plots to visualize relationships between features.

5. Used histograms to see the distribution of each feature.

6. Used box plots to identify potential outliers.

5. Visual Results & Interpretation

- Scatter plots showed clear separation among species, especially with petal measurements.

- Histograms revealed the normal-like distribution for some features.

- Box plots helped detect outliers in sepal width.

- These visualizations helped us quickly understand patterns in the data.

6. Final Insights

The Iris dataset is well-structured and suitable for visualization. Most features show clear distribution patterns, and species can be visually separated based on petal dimensions. This makes it an ideal dataset for beginner-level data analysis tasks.