

Exploratory Data Analysis (EDA) Report

Dataset: Iris Flower Dataset

Objective: Understand data distribution, relationships, outliers, and important features for prediction.

1. Histogram Analysis

- 1 Sepal length and petal length show clear variation among samples.
- 2 Petal features have better separation compared to sepal features.
- 3 No heavy skewness observed — data is well balanced.
- 4 Insight: Petal dimensions are more informative for classification.

2. Categorical Count Plot

- 1 Each species (Setosa, Versicolor, Virginica) has equal samples.
- 2 No class imbalance.
- 3 Insight: Balanced dataset is good for training ML models.

3. Box Plot (Outlier Detection)

- 1 Minor outliers observed in sepal width.
- 2 No extreme anomalies.
- 3 Insight: Dataset is clean and reliable — minimal preprocessing required.

4. Correlation Heatmap

- 1 Strong positive correlation between petal length and petal width.
- 2 Strong positive correlation between sepal length and petal length.
- 3 Weak correlation observed for sepal width.
- 4 Insight: Petal features dominate prediction power.

5. Important Features for Prediction

- 1 Petal Length
- 2 Petal Width
- 3 Sepal Width (less predictive)

6. Summary Findings

- 1 Data is balanced and clean.
- 2 Strong feature relationships exist.
- 3 Petal features are the most predictive.
- 4 Minimal outliers.
- 5 Visualization clearly reveals class separation.