IRIS classification

Name: Khushi Rajpurohit

Problem Statement: Classify the IRIS dataset for future prediction

Dataset used: I used IRIS dataset from kaggle.

Data exploration: head(),.shape(),.info(),.describe(), .duplicated() and .isnull().sum() methods are used to get an overview of dataset, its structure and missing values.

Exploratory Data Analysis:

- Visualizing the count between all species
- Correlation between different features.

Insights from Analysis:

- All the species are of same size. (i.e 50)
- High correlation between sepal length with species, petal length and petal width.
- High correlation between petal length and species.
- High correlation between petal width and species.

Predictive Modelling:

- **Prepared data For Modelling**
    - **Feature Selection**
    - **Encoding categorical variable to numerical**
    - **Splits data into training set and testing set**
    - **Converted the values to array for easy prediction**

- Trained a logistic regression, KNN, decision Tree model and evaluated its performance using metrics like accuracy score.
- Logistic Regression:
    - Accuracy: 100%
- Decision Tree:
    - Accuracy:97%
- KNN:
    - Accuracy:100%

Conclusion: we can either choose Logistic Regression or KNN for further prediction