NEXUS PROJECT -1
DATA SCIENCE
Iris Dataset Classification Project Documentation

Introduction:

This project involves the use of machine learning techniques to classify iris flowers based on their features. The dataset used is the famous Iris dataset, which contains measurements of sepal length, sepal width, petal length, and petal width for three species of iris flowers (setosa, versicolor, and virginica).

1. **Importing Libraries:**

The necessary libraries such as Pandas, NumPy, Matplotlib, Seaborn, and scikit-learn are imported.

2. **Loading the Dataset:**

The Iris dataset is loaded using scikit-learn's `load_iris` function. The data is then converted into a Pandas DataFrame for easy manipulation.

3. **Exploratory Data Analysis (EDA):**

- Basic statistics of the dataset are displayed using `describe()`.
- The distribution of each feature is visualized using histograms.
- Relationships between features are explored through pairplots.

4. **Data Splitting:**

The dataset is split into training and testing sets using scikit-learn's `train_test_split`function.

5. **Model Selection and Training:**

- Logistic Regression is chosen as the classification model.
- The model is trained on the training set using `LogisticRegression` from scikit-learn.

6. **Model Evaluation:**

- The trained model is used to predict the target values on the testing set.
- Performance metrics such as accuracy, precision, and recall are calculated using scikit-learn's metrics functions.

7. **Results:**

The evaluation metrics (accuracy, precision, and recall) are printed to assess the performance of the trained model on the testing set.