TASK 1: Iris Flower Classification Using Machine Learning

Step 1: Problem Understanding

The objective of this task is to classify Iris flowers into three species: Setosa, Versicolor, and Virginica. The classification is based on four flower measurements: sepal length, sepal width, petal length, and petal width. This is a supervised learning classification problem because the target variable is categorical.

Step 2: Import Required Libraries

Python libraries such as Pandas, NumPy, Matplotlib, and Scikit-learn are imported. These libraries help in data loading, analysis, visualization, model training, and evaluation.

Step 3: Load the Dataset

The Iris dataset is loaded using Pandas. The dataset contains flower measurements along with their corresponding species labels.

Step 4: Exploratory Data Analysis (EDA)

EDA is performed to understand the structure and quality of the dataset. This includes checking the first few rows, dataset size, data types, and class distribution. EDA helps identify patterns, balance between classes, and any data issues.

Step 5: Data Cleaning

Unnecessary columns such as the ID column are removed from the dataset. Data cleaning ensures that only relevant features are used for model training, improving performance.

Step 6: Feature Selection

The dataset is divided into independent variables (features) and the dependent variable (target). Features include all measurement columns, while the target variable is the flower species.

Step 7: Train-Test Split

The dataset is split into training and testing sets. The training set is used to train the model, while the testing set evaluates its performance. Typically, 80 percent of the data is used for training and 20 percent for testing.

Step 8: Model Selection

Logistic Regression is chosen as the machine learning model. It is suitable for multi-class classification problems, simple to implement, and easy to interpret.

Step 9: Model Training

The Logistic Regression model is trained using the training data. During training, the model learns the relationship between flower measurements and species labels.

Step 10: Prediction and Evaluation

The trained model predicts flower species for the test dataset. Model performance is evaluated using accuracy score and classification report. High accuracy indicates good classification performance.

Step 11: Visualization

Visualization is used to understand how different Iris species are distributed based on features. Scatter plots of sepal and petal measurements clearly show separation between species. Petal features show strong distinction, explaining high model accuracy.

Step 12: Result Interpretation and Conclusion

The Iris Flower Classification model successfully classifies flowers into their respective species. The results show that petal measurements play a key role in species classification. This model demonstrates the effectiveness of supervised machine learning in classification tasks.