Iris Flower Classification using Machine Learning

1. Introduction

Machine learning is widely used in classification problems where the objective is to categorize data into predefined classes. The Iris flower classification problem is a classic machine learning task that involves predicting the species of a flower based on its physical characteristics such as sepal length, sepal width, petal length, and petal width.

This project demonstrates the use of supervised machine learning techniques to classify iris flowers into different species.

---

2. Problem Statement

The objective of this project is to build a machine learning model that can accurately classify iris flowers into one of the three species — Iris-setosa, Iris-versicolor, or Iris-virginica — based on their sepal and petal measurements.

---

3. Dataset Description

The Iris dataset contains 150 records and 6 columns:

- Id
- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species

The Id column does not contribute to classification and is removed during preprocessing. The Species column represents the target variable.

---

4. Data Preprocessing

Data preprocessing involves the following steps:

- Removing the Id column as it has no predictive value
- Separating features (X) and target labels (y)
- Splitting the dataset into training and testing sets using an 80:20 ratio

This ensures that the model is trained on one part of the data and evaluated on unseen data.

---

5. Model Selection

The **K-Nearest Neighbors (KNN)** algorithm is used for classification. KNN works by identifying the nearest data points to a given input and assigning the class based on majority voting.

KNN is suitable for this problem because:

- The dataset is small
- Features are numerical
- Classes are well-separated

---

6. Model Training

The KNN classifier is trained using the training dataset. The value of k (number of neighbors) is set to 1 to achieve high classification accuracy for this dataset.

The trained model learns the distance patterns between different flower species.

---

7. Prediction and Testing

After training, the model is tested using unseen data. Additionally, a custom input containing sepal and petal measurements is provided to the model to predict the flower species.

The model successfully predicts the species as **Iris-setosa** for the given input values.

---

8. Results

The model performs accurately in classifying iris flowers. The prediction results confirm that the trained KNN classifier can correctly identify flower species based on physical measurements.

---

9. Conclusion

In this project, a machine learning model was developed to classify iris flowers using the K-Nearest Neighbors algorithm. Proper preprocessing and feature selection played a key role

in achieving accurate predictions. The project demonstrates the effectiveness of supervised learning for classification tasks.

---

10. Future Scope

The model can be improved by:

- Experimenting with different values of k
- Using other classification algorithms such as Logistic Regression or Support Vector Machines
- Applying cross-validation for better generalization