MINOR PROJECT ON IRIS DATASET

To perform classification analysis on Iris dataset. Perform any two classification algorithms and compare the accuracy.

ABOUT IRIS DATASET:

The iris dataset contains the following data:

- 50 samples of 3 different species of iris (150 samples total)
- Measurements: Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, Petal WidthCm
- The format for the data: (Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, Petal WidthCm)
- The variables are:

- SepalLengthCm: Sepal length, in centimeters, used as input.
- SepalWidthCm: Sepal width, in centimeters, used as input.
- PetalLengthCm: Petal length, in centimeters, used as input.
- PetalWidthCm: Petal width, in centimeters, used as input.
- Species: Iris Setosa, Versicolor, or Virginica, used as the target.

CONTENTS

1. TOOLS I HAVE USED

Jupyter notebook for implementation of codes.

2.DATA PREPROCESSESING

- Include Libraries: Import Libraries such as pandas, numpy, matplotlib, seaborn and some packages from scikit-learn.
- Import Dataset: Import the dataset IRIS
- Handle the Missing Values: Check whether there are any missing values in the dataset.

3.DATA VISUALIZATION

- **Scatterplot**
- **Pairplot**
- **Boxplot**
- **Correlation**

4. FEATURE ENGINEERING

DIVIDING THE DATA INTO FEATURES AND LABELS

As we can see dataset contain six columns: Id, SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm and Species. The actual features are described by columns 1-4. Last column contains labels of samples. Firstly, we need to split data into two arrays: X (features) and y (labels).

5.MACHINE LEARNING ALGORITHMS

I trained my model using several Machine Learning Algorithms and compared their results. The Machine Learning Classification Algorithms which I implemented are:

1.DECISION TREE ALGORITHM: Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

2.SUPPORT VECTOR MACHINE ALGORITHM: "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems.

3. K-NEAREST NEIGHBOUR ALGORITHM: "K-NN algorithm" assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity.

OBSERVATIONS AND RESULTS:

By comparing the accuracy_score of the mentioned algorithms,

The algorithms which gave more accuracy are:

Decision Tree → 1.0
Support Vector Machine →0.966666667
K-Nearest Neighbours →0.96666666667

CONCLUSION:

By Using these models I have trained using Machine Learning Algorithms we c an predict the species of the Iris flower, whether it is an 'Iris-setosa' , 'Iris-virginica', Iris-versicolor'.

PROJECT COMPLETED BY

K. UMA PRANAVI

ML-MINOR-JUNE