

Verzeo Minor Project Submission

To perform classification analysis on IRIS dataset

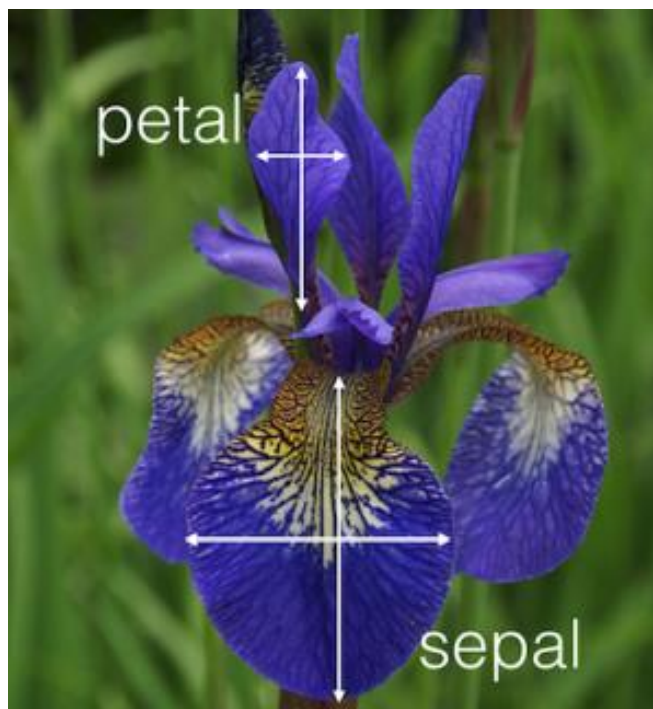
By- Suryaansh Jaiswal

Project review

I have made a ML project to demonstrate classification on the IRIS dataset.

The Iris dataset can be used by a machine learning model to illustrate classification. Classification is a type of supervised machine learning problem where the target variable is categorical. It is a method used to determine the type of an object by comparison with similar objects that have previously been categorised. Once trained on known data, the machine learning model can make a predictive classification by comparing a test object to the output of its training data..

The IRIS dataset was generated in 1936 by the British statistician and biologist Ronald Fisher. It contains 150 samples in total, comprising 50 samples of 3 different species of Iris plant (Iris Setosa, Iris Versicolour and Iris Virginica). For each sample, the flower measurements are recorded for the sepal length, sepal width, petal length and petal width, as illustrated in the following graphic.



Tools Used

Python- Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation.

Google colab- Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX and more.

Iris Dataset- The Iris dataset can be used by a machine learning model to illustrate classification

Algorithm

KNN-

The k-nearest neighbors (KNN) algorithm is used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slower as the size of that data in use grows.

It is one of the fundamental algorithms in **machine learning**. Machine learning models use a set of input values to predict output values. KNN is one of the simplest forms of machine learning algorithms mostly used for classification. It classifies the data point on how its neighbor is classified.

Decision Tree-

Decision Trees are **a type of Supervised Machine Learning** (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves.

A decision tree is **a very specific type of probability tree that enables you to make a decision about some kind of process**. For example, you might want to choose between manufacturing item A or item B, or investing in choice 1, choice 2, or choice 3.

Conclusion-

The accuracy calculated after using KNN is 1

Accuracy

```
[ ] from sklearn.metrics import accuracy_score
```

```
acc= accuracy_score(y_test,y_pred)  
acc
```

```
1.0
```

Hence the accuracy from KNN comes out to be 1

The accuracy calculated after using Decision Tree is 0.966

Accuracy and Confusion Matrix

```
from sklearn.metrics import confusion_matrix , accuracy_score
```

```
[20] confusion_matrix(y_test,y_pred)
```

```
array([[24,  0],  
       [ 1,  5]])
```

```
accuracy_score(y_test,y_pred)
```

```
0.9666666666666667
```

Hence the accuracy from Decision Tree is 0.966

Hence the accuracy calculated by KNN is more

Thank You
