
The iris Dataset Analysis(Report)

prepared By -Rushika Nikam

Github link= <https://github.com/Rushikanikam15>

Information About dataset :-

In this data will do some exploratory data analysis on the Iris dataset. The Iris Dataset contains four features (length and width of sepals and petals) of 50 samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor), this data Classify iris plants into three classes .dataset is often used in data mining, classification examples and to test the algorithms.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   sepal_length    150 non-null    float64
 1   sepal_width     150 non-null    float64
 2   petal_length    150 non-null    float64
 3   petal_width     150 non-null    float64
 4   class           150 non-null    object  
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

Just for reference, here are pictures of the three flower class



Iris Versicolor



Iris Setosa



Iris Virginica

Exploratory Data Analysis for iris data:-

- ❖ For using pairplot We can see many types of relationships from this plot such as the class setosa has the smallest of petals widths and lengths. It also has the smallest sepal length but larger sepal widths. Such information can be gathered about any other class.
- ❖ Comparing petal length and petal width:-Class setosa has smaller petal lengths and widths, versicolor class lies in the middle of the other two classes in terms of petal length and width Class virginica has the largest of petal lengths and widths.
- ❖ Comparing sepal length and sepal width:-class setosa has smaller sepal lengths but larger sepal widths. Versicolor class lies in the middle of the other two class in terms of sepal length and width, Class virginica has larger sepal lengths but smaller sepal widths .the Most the columns is continuous in natures except Class which is object datatype.
- ❖ By observing the descriptive statistics of the data, we can say that Petal length column has little variance in it, rest every other column is almost normally distributed.
- ❖ Skewness present in data
- ❖ Petal Length and Petal width are the most useful attributes to identify flowers types. Being linearly separable, Setosa can be easily identified, virginica and versicolor have some overlap

Data sampling[Splitting Data in train test part]

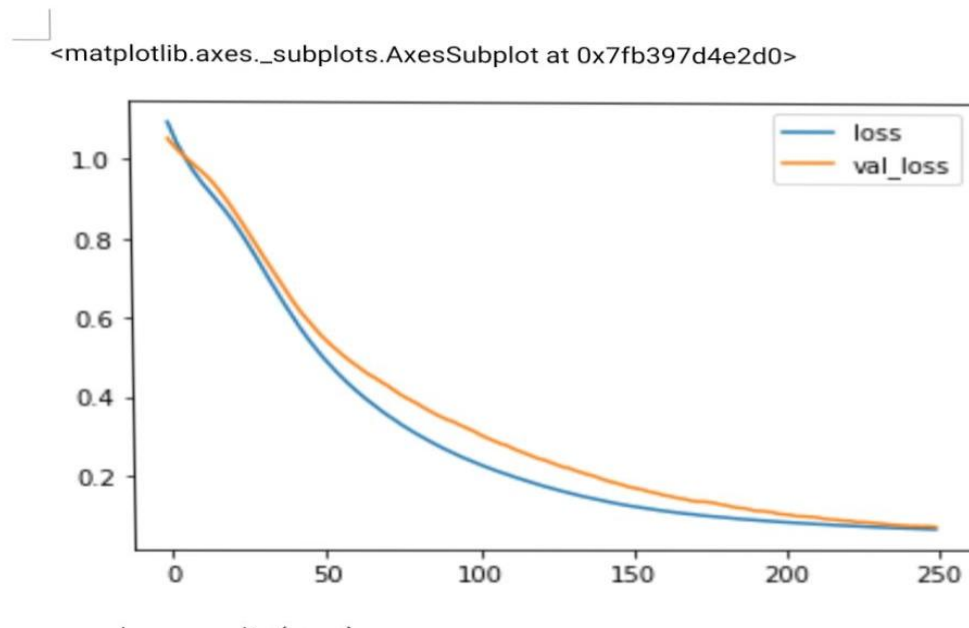
- ❖ In the data-sampling step, a subset of data is extracted from the input dataset. This is performed for sampling and for dividing the data into the two classes of training and testing data.
- ❖ I have kept test size is 0.20 I.e I am sending 20% of the data for testing and remaining 80% data for training.
- ❖ I have Build artificial Neural Network model on that data.

Algorithm

- ❖ In this step, the parameters of the algorithms are tested and evaluated by using different values in order to find the best possible result in the system output
- ❖ I have used artificial neural network model for this data
- ❖ This kernel uses perceptrons (Neural Network) to predict the class of the Iris dataset. Neural network is a machine learning algorithm which is inspired by a neuron.
- ❖ A neuron consists of a dendrite and an axon which are responsible for collecting and sending signals. For our artificial neural network, the concept works similar in which a lot of neurons are connected to each layer with its own corresponding weight and biases.
- ❖ A perceptron cannot read strings, the variety's names are mapped to integers, Iris Setosa is assigned to 0, Iris Versicolor is assigned to 1 and Iris Virginica is assigned to 2

- ❖ Although there are currently architecture of neural network, perceptron is being used as the architecture to prevent overfitting to the Iris classes due to less feature.

Loss and validation_loss:-



Training results for IRIS dataset. a Variation of train loss/ error with number of epochs.

Conclusion:-

Yes, you can classify the new flower belonging to one of three classes (Setosa, Versicolor and Virginia) using iris Data set by applying various data analysis techniques and using various Python libraries.

```
from sklearn.metrics import classification_report
print(classification_report(ytest,ypred))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	11
1	1.00	0.92	0.96	13
2	0.86	1.00	0.92	6
micro avg	0.97	0.97	0.97	30
macro avg	0.95	0.97	0.96	30
weighted avg	0.97	0.97	0.97	30
samples avg	0.97	0.97	0.97	30

```
from sklearn.metrics import accuracy_score
print(accuracy_score(ytest,ypred))
```

```
0.9666666666666667
```

Got a 96% accuracy by using artificial neural network model