

Analysis and Visualization of an Iris data set

The Iris bloom informational collection or Fisher's Iris informational index is perhaps the most popular multivariate datum set utilized for testing different Machine Learning Algorithms. This is my variant of EDA on Iris Dataset. Information experiences of every single perception step has been given

Information Visualization : It is a method for communicating your information in a visual setting so that examples, connections, patterns between the information can be handily perceived. Information Visualization helps in tracking down secret experiences by giving skin to your crude information (skeleton).In this task, we will utilize different datasets to show precisely the way that things work. The base dataset will be the iris dataset which we will import from sklearn. We will make the remainder of the dataset as indicated by need.

How about we import every one of the libraries which are expected for doing When you complete this ability way, you'll be prepared for further developed information expert courses. Draw in with our well disposed local area of information examiners, get input on your activities, and continue to assemble your abilities.

Machine Learning you will be practicing basic ML algorithms on this famous dataset. Iris dataset contains five columns such as Petal Length, Petal Width, Sepal Length, Sepal Width and Species Type.

Reading and analyzing data set

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

Iris dataset contains five columns such as Petal Length, Petal Width, Sepal Length, Sepal Width and Species Type.

Iris is a flowering plant, the researchers have measured various features of the different iris flowers and recorded digitally.

Statistics of data set

Some of the data insights are

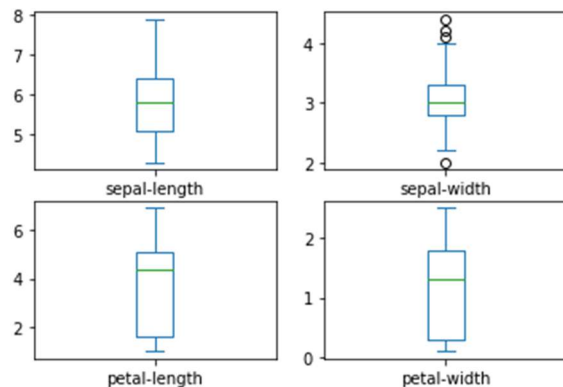
- 1 All columns are not having any Null Entries
- 2 Four columns are numerical type
- 3 Only Single column categorical type

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

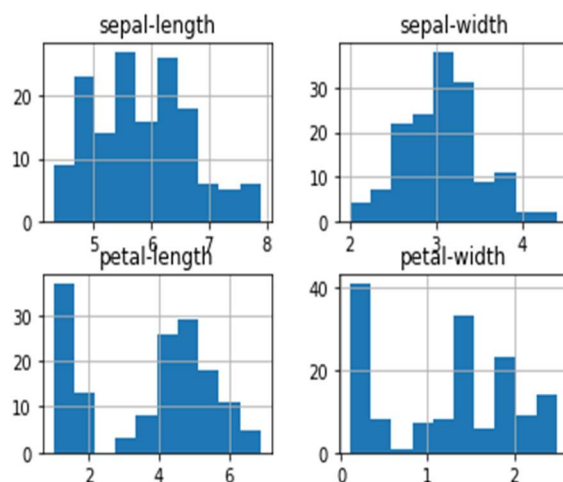
We can see that only one column has categorical data and all the other columns are of the numeric type with non-Null entries.

Let's get a quick statistical summary of the dataset using the `describe()` method. The `describe()` function applies basic statistical computations on the dataset like extreme values, count of data points standard deviation, etc. Any missing value or Nan value is automatically skipped.

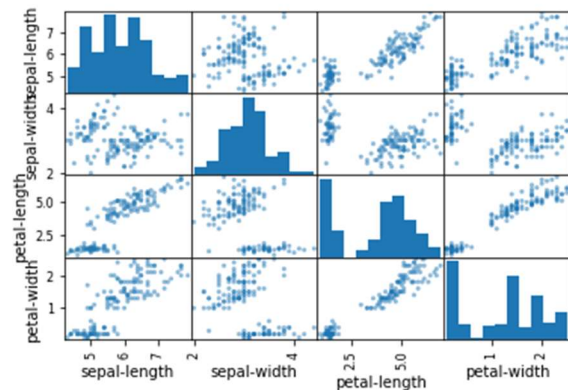
Plots:



Above Box-whisker plot shows difference in the data sets. And gives additional details of data plot. Setosa species has smaller sepal length but higher width. Versicolor lies in almost middle for length as well as width. Virginica has larger sepal lengths and smaller sepal widths.



The above Histogram graph shows the difference of sepal-length, sepal-width, petal-length, petal-width. The Setosa species has smaller sepal length but higher width. Versicolor lies in almost middle for length as well as width. Virginica has larger sepal-length and smaller sepal widths.



The above scatter plot shows it is clear that Setosa species have the smallest petal length as well as petal width. Versicolor species have average petal length and petal width. Virginica species have the highest petal length as well as petal width.