Hi Everyone,

My Name is **Nivitus.** Welcome to the Iris Flower Classification Tutorial. This is my first Machine learning blog on Medium Site. I hope all of you like this blog; ok let’s go to the topic. Today

We are going to implement the model which is Iris Classification, The problem is going to do solve by one of the Algorithms on Machine Learning which is called SVM (Support Vector Machine).

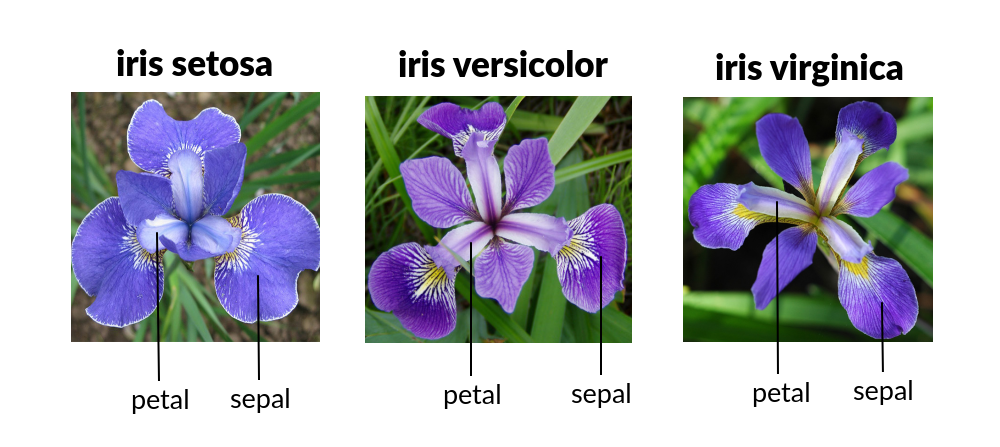
Let’s get started.

**A Small Introduction about Iris Dataset**

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician, eugenicist, and biologist Ronald Fisher in his 1936 paper the use of multiple measurements in taxonomic problems as an example of linear discriminant analysis.

His contribution to statistics is way beyond the Fisher’s exact test. For example, he developed the maximum likelihood estimation and the analysis of variance (more commonly known as its acronym ANOVA) test. For these important contributions, he has been highly regarded in the history of modern statistics, as noted on his Wikipedia page.

**Iris Flower**



All this aside, we know that the three Iris species in the dataset: Iris setosa*,*Iris virginica*, and*Iris versicolor*,* and we also know that the dataset records the lengths and widths of sepals and petals for these flowers.

**Our Goal:**

Create the Machine Learning model that can classify the different species of the Iris flower.

**Steps towards the Good Model:**

* Make dataset
* Do some Visualizations(EDA)
* Build the model,
* Train the model.
* Make predictions.

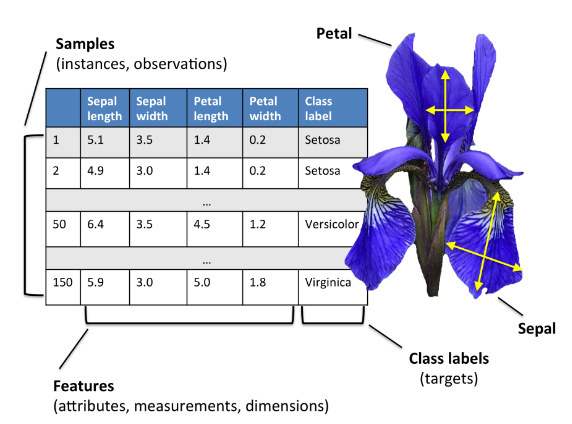
1. **Make Dataset**

The Iris Dataset is already available in the ***sklearn*** library we just have the ***datasets.load\_iris()***  code in Python. In Other hand we can able to do download the ***Kaggle*** also.

Come on let’s make a understanding this one

**The data set consists of:**

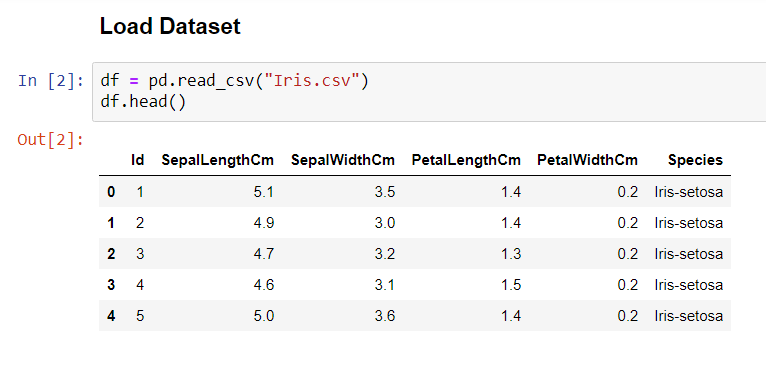
* 150 data rows
* 3 Prediction labels: species of Iris (*Iris setosa, Iris virginica* and *Iris versicolor*)
* 4 features: Sepal length, Sepal width, Petal length, Petal Width.



Scikit learn only works if data is stored as numeric data, irrespective of it being a regression or a classification problem. It also requires the arrays to be stored at numpy arrays for optimization.

1. **Do Some Visualizations (EDA)**

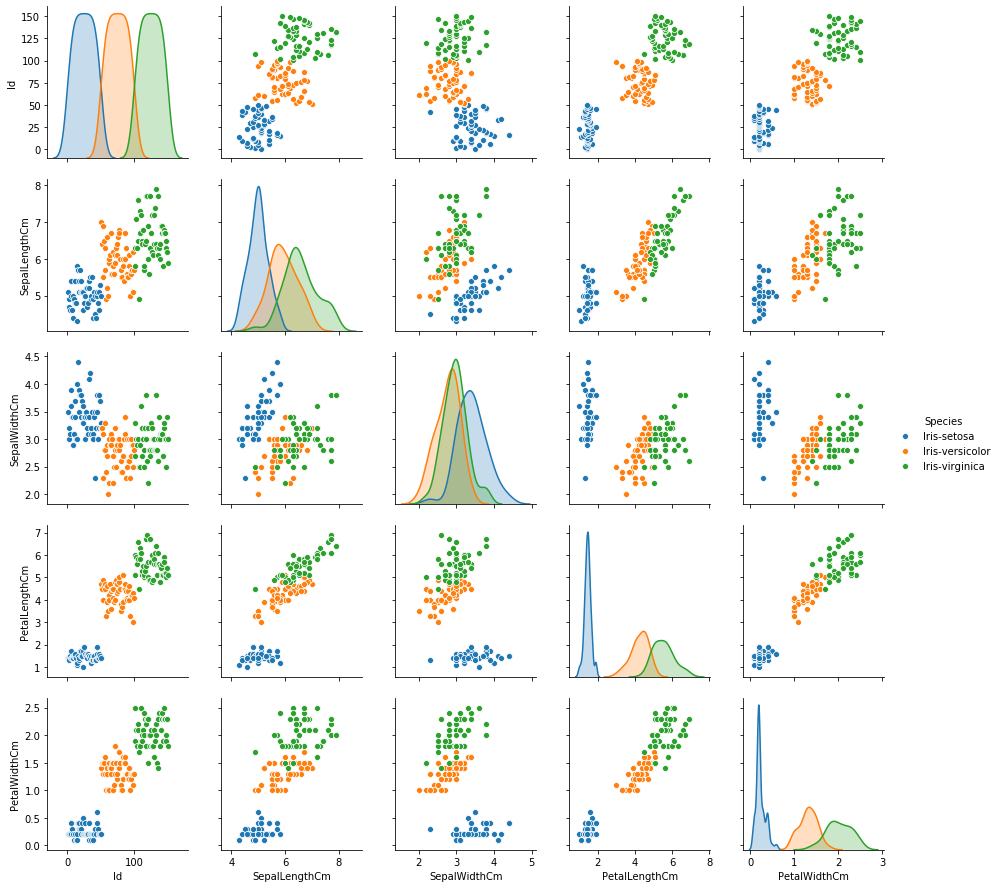
In statistics, exploratory data analysis (**EDA**) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily **EDA** is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.



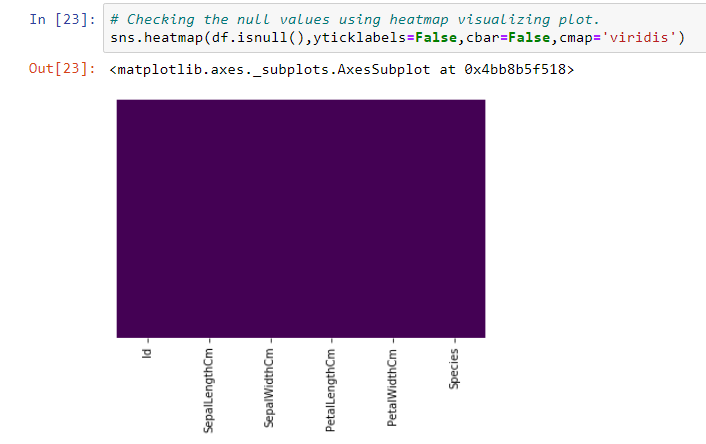
After loading the dataset using csv file format the dataset will be look like.

Code: **sns.pairplot(df, hue='Species')**

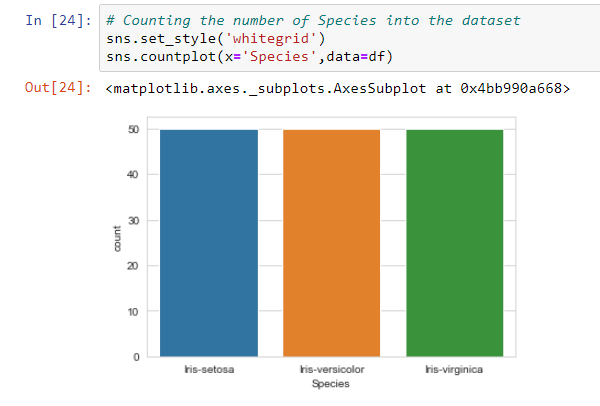
Here sns Alice for the seaborn library which really help to visualization data part in learning.



**Checking the null values using heatmap**



**Counting the data points using Box Plot**



**3. Train Test Split and Build and Train the model**

Since our process involves training and testing, we should split our dataset. It can be executed by the following code.

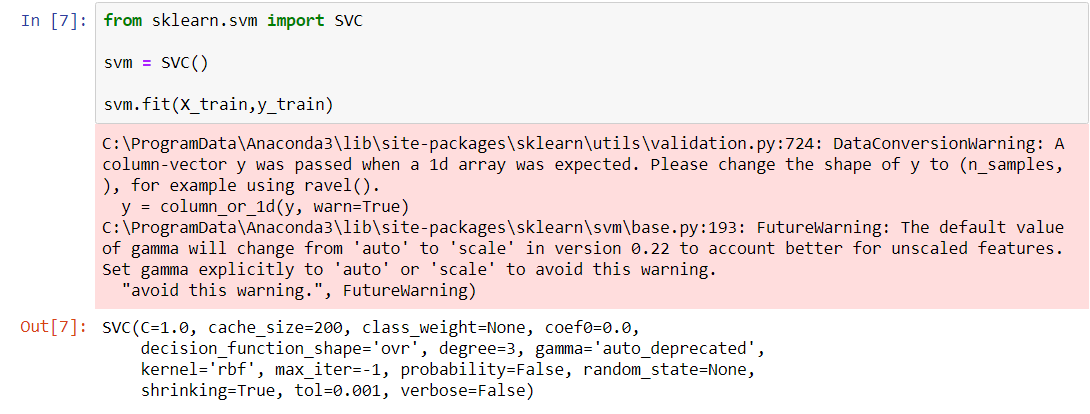


X\_train has the training features

X\_test has the testing features

y\_train has the training label

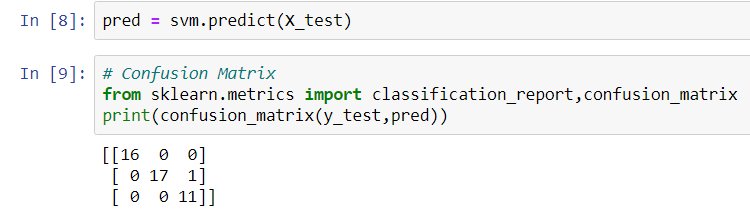
y\_test has the testing labels

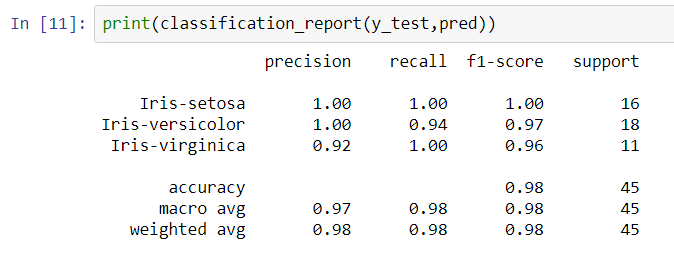


Just now we fit the model using sklearn and fit method. Here svm is the classifier for SVC method.

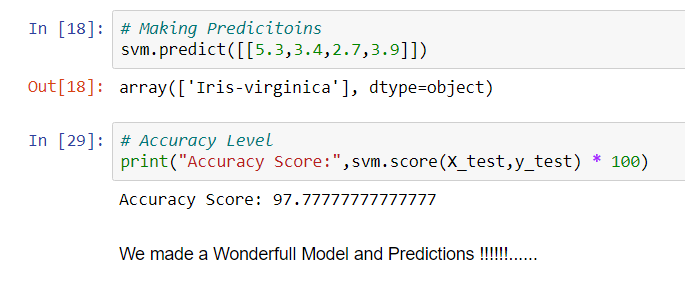
**4. Predict the Model**

Predictions done with the help of predict method in sklearn. As well lets we use the confusion matrix to find the prediction.





This is a classification report for the Iris Flower Dataset such as precision, recall, f1-score, and support.



**Output:**

Our Final Prediction score is 97.7777

We made it!!!!!! ☺☺☺

Keep on Learning Never Give up… ☺☺☺

