**Name:Dipali lasurkar**

**Class :BE comp (B).**

**ML: Mini project No. 1**

**Title:** Apply the Genetic Algorithm for optimization on a dataset obtained from UCI ML repository. (IRIS Dataset ).

**Prerequisite:**

-Basic of Python, Data Mining Algorithm, Iris Dataset ,Genetic algorithm.

**Software Requirements:**

-Anaconda with Python 3.7

**Hardware Requirement:**

-PIV, 2GB RAM, 500 GB HDD, Lenovo A13-4089Model.

**Learning Objectives:**

-Learn How to Apply Genetic Algorithm for given Iris Dataset. The main objective of this assignment is to implement Iris Flower Dataset or any other dataset into a data frame using python .

**Outcomes:**

-After completion of this assignment students are able Implement code for the Iris Dataset with plotting diagram.

**Theory Concepts:**

**1.** Python is an interpreted high level programming language for general purpose programming created by Guido Van Rassom and First released in 1991.

**2.** Python for a design philophy that emphasizes code readability, notably using significant white space .

**3.** Python features a dynamic type of automatic memory management support multiple programming paradigm ,including object – oriented, imperative ,functional and procedural and has a large ,comprehensive standard library.

**4.** Python library is a collection of function and methods that allows you to performs lots of actions without writing your own code.

Eg: If you are working with data, numpy , scipy, pandas ,etc .are the libraries you must know.

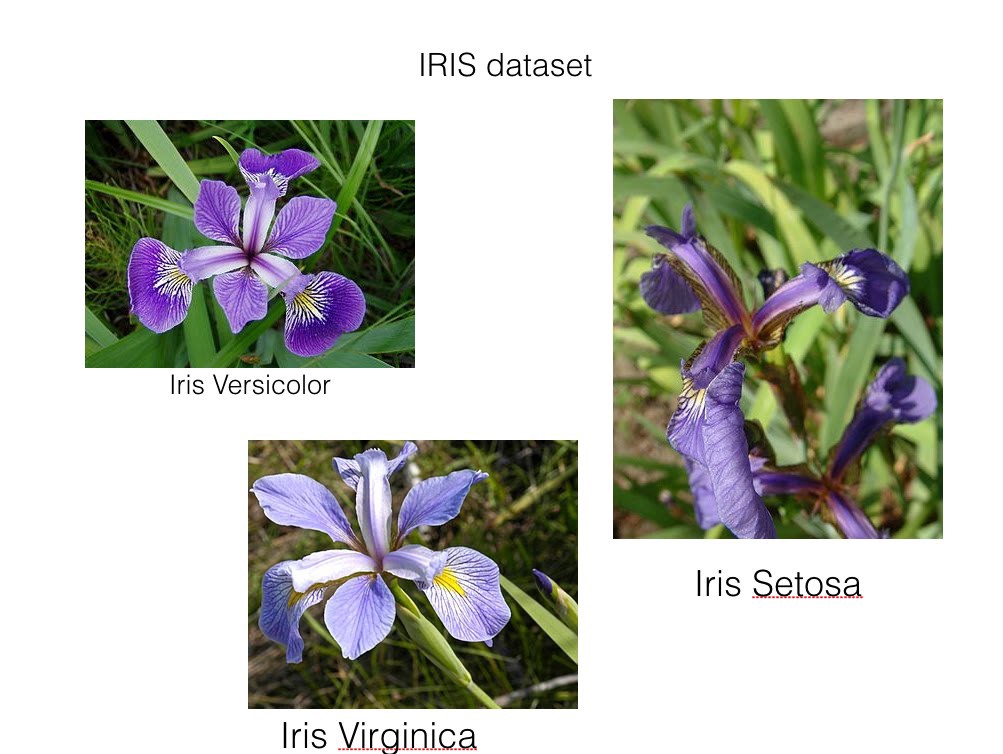
* **Import pandas as pd**

Pandas is an open source ,BSD-licensed library providing high performance, easy to use data structure and data analysis tools for the python programming language.

* sudo apt-get install python 3.6
* sudo apt-get install python pip
* sudo apt-get install python pandas
* **Import matplotlib as plt**
* Matplotlib is a plotting library for the python programming language and its numerical mathematics extension numpy.
* If provides an object oriented API for embedding plot into applications using general purpose GUI tools like Tkinter , Wxpython.
* **Iris Dataset**
* This dataset includes three species with so samples each as well as some properties about each flower .
* The available columns in this dataset are : id , sepal length cm, sepal width cm, petal length cm , petal width cm and species .
* The Dataset is self available below in csv file . This dataset is also available in scikit-learn package of which the link description also attached in title.
* The main task in this dataset is to create an iris (name of a flower)

Classifier based on given properties that are the sepal and petal size.

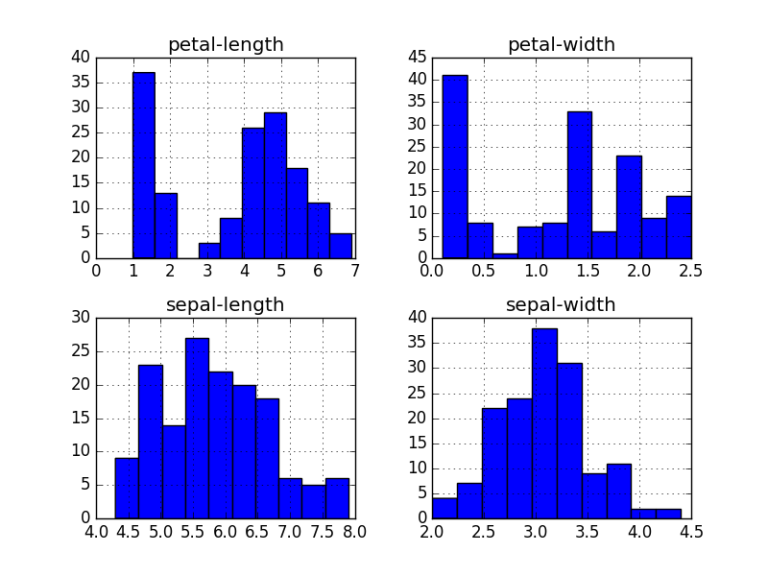
* If you don’t know the difference between sepal and petal, here is an image that shows which part of the flower is sepal and which part is petal.



* df.isnull().any()
* It is used to check whether we have null values in our dataset or not .
* df.types()
* To know the type of each column values .
* df.describe()
* check the quick summary of data.
* Df.[‘petalwidth’].plot.hist()

Plt.show()

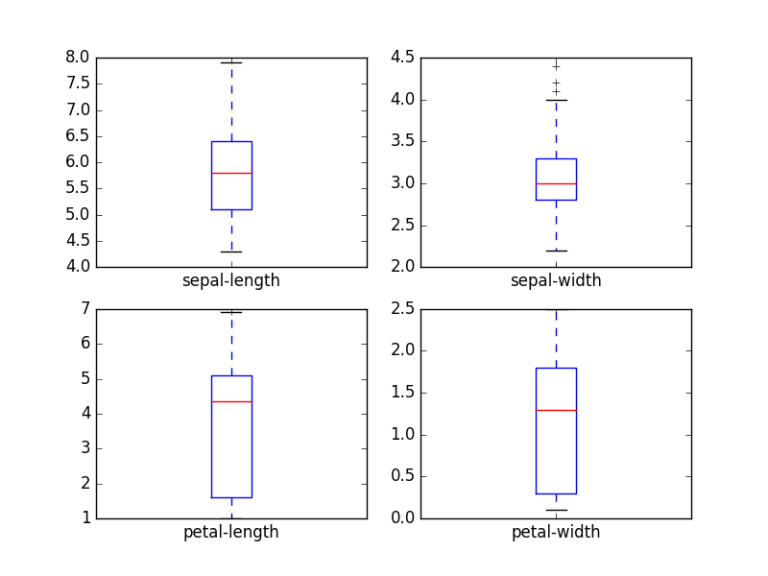
* It is used to represent flowers datasets of values between 0.1 and 0.5 in the graph form.



* **Splitting the Dataset**
* Since there is only dataset available, we need to divide the dataset into training and test dataset.
* To do this, we can use train\_test\_split method from the scikit learn.
* **BoxPlot**
* A ‘Boxplot’ or ‘box-&-whiskerplot’ is a graphical summary of the distribute.
* The box in the middle indicates ‘hinges’ and ‘median’.
* The lines(‘whisker’)show the largest or smallest observation that falls within a distance of 1.5 times the box size from the nearest hinge.

If any observation fall farther away, the additional points are considered ‘extreme’ values and are shown separately.

* A boxplot can often give a good idea of the data distribution and is often more useful to compare distributions side by side as it is more compact than histogram.
* We can use the boxplot function to calculate quick summaries for all the variables in our dataset by default.
* The real power of boxplots is really to do comparisons of variables by sub-grouping.



## Conclusion

## Hence ,We have studied and practically implemented Iris flower dataset into a Data frame And we learn Genetic Algorithm for optimizing Iris Dataset .