**TABLE OF CONTENTS**

**PAGE NO**

**LIST OF FIGURES**

1. **Pair plot 4**
2. **Sepal\_Length vs Petal\_Length 5**
3. **Sepal\_Width vs Petal\_Width 5**
4. **Axes plot 5**
5. **Axes grid 6**
6. **Fig 7**
7. **Box plot 7**
8. **Heat map 8**
9. **Predictions for Species 8**

**SECTION 1:** **INTRODUCTION TO THE STUDY**

It includes three iris species with 50 samples each as well as some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.

**The columns in this dataset are:**

Id

sepal\_length

sepal\_width

petal\_length

petal\_width

species

**Intro to Python Data Visualization and Data Science using Iris Data:**

Simple project for visualizing iris species dataset and identifying Iris flower species based on its sepal and petal dimensions. In this notebook we'll play with IRIS data, an introductory dataset for DS Learners. Here we will plot Iris data using python and will build some DS models. Get your hands dirty!

Data Set Information:

This is perhaps the best-known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

**Dependencies**

This project requires Python 3.5 and the following Python libraries installed:

Scipy

Scikit Learn

NumPy

Matplotlib.Pyplot

Seaborn

You will also need to have software installed to run and execute jupyter notebook.

Install anaconda, a pre-packaged Python distribution that contains all of the necessary libraries and software for this project.

Models: Finding Best Accuracy Model

Best accuracy found on Support Vector machine - 0.95

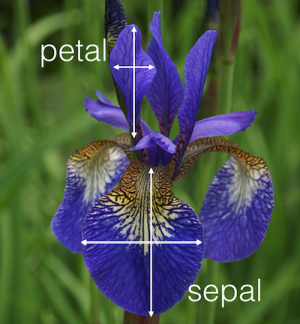
**Requirements for working with datasets in scikit-learn**

Additional resources

The iris dataset contains the following data.50 samples of 3 different species of iris (150 samples total)

Measurements: sepal length, sepal width, petal length, petal width

The format for the data: (sepal length, sepal width, petal length, petal width)



**Data Science terminology**

Each row is an observation (also known as: sample, example, instance, record)

Each column is a feature (also known as: predictor, attribute, independent variable, input, regressor, covariate)

**Exploring the iris dataset**

Each value we are predicting is the response (also known as: target, outcome, label, dependent variable)

Classification is supervised learning in which the response is categorical

"0": setosa

"1": versicolor

"2": virginica

Regression is supervised learning in which the response is ordered and continuous any number (continuous)

Requirements for working with data in scikit-learn Features and response are separate objects. In this case, data and target are separate. Features and response should be numeric. In this case, features and response are numeric with the matrix dimension of 150 x. Features and response should be NumPy arrays. The iris dataset contains NumPy arrays already. For another dataset, by loading them into NumPy

Features and response should have specific shapes

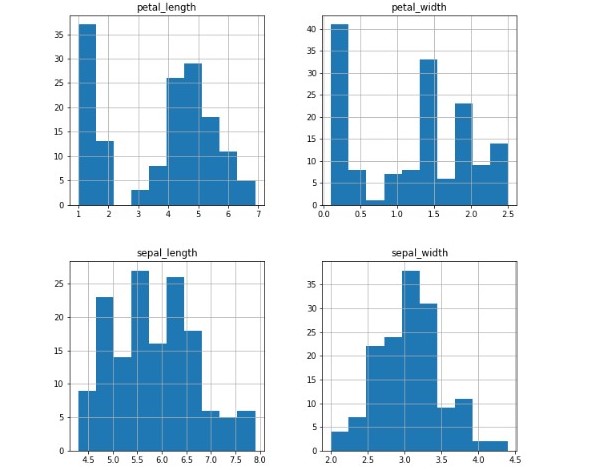
150 x 4 for whole dataset

150 x 1 for examples

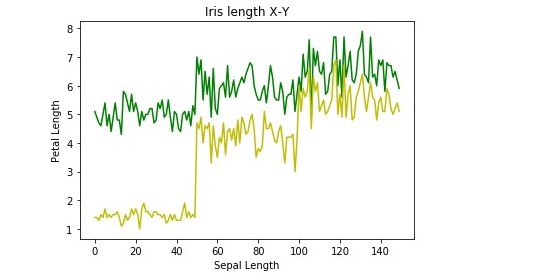
4 x 1 for features

you can convert the matrix accordingly using np.tile(a, [4, 1]), where a is the matrix and [4, 1] is the intended matrix dimensionality

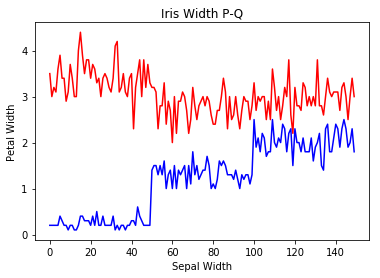
**GRAPHS:**



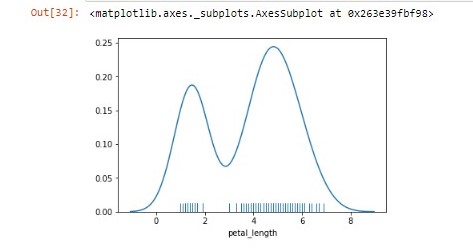
**Graph of different sepal\_length and sepal\_width pair plot**



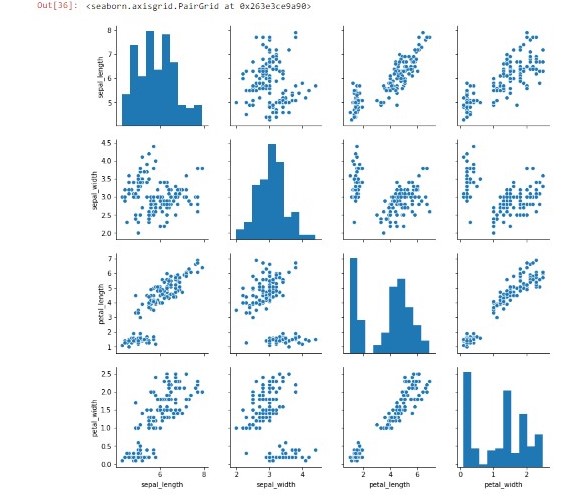
**Sepal\_Length vs Petal\_Length**

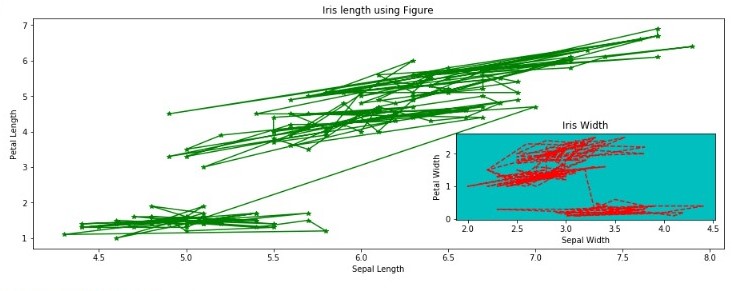


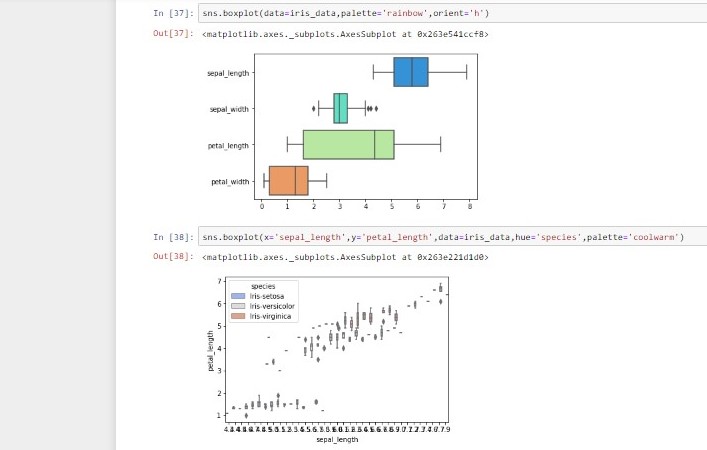
**Sepal Width vs Petal Width**



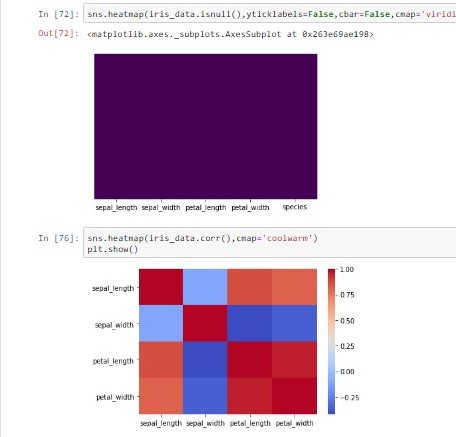
**Axes plot**

**SEABORN AXISGRID**

**Iris length using fig**



**BOX PLOT**



**SEABORN USING HEATMAPS**

A screenshot of a social media post

Description automatically generated

**Prediction Table For Species**

* These are some graphs used in my project. Also, they are still some methods used in my project which are not mentioned in the Report such as Seaborn different model Graphs and predicted the prediction table to find the approximate species.