Assignment IRIS Dataset

Table of Contents

[About Dataset and Algorithm: - 2](#_Toc65095534)

[Table of Observation 2](#_Toc65095535)

[Python Code 3](#_Toc65095536)

[Scatterplot for K = 12 5](#_Toc65095537)

Name: - Vikash Singh

Roll No: - 100150

VIT Bhopal

# About Dataset and Algorithm: -

The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

This data sets consists of 3 different types of irises (Setosa, Versicolour, and Virginica) petal and sepal length, stored in a 150x4 numpy.ndarray. The rows being the samples and the columns being: Sepal Length, Sepal Width, Petal Length and Petal Width.

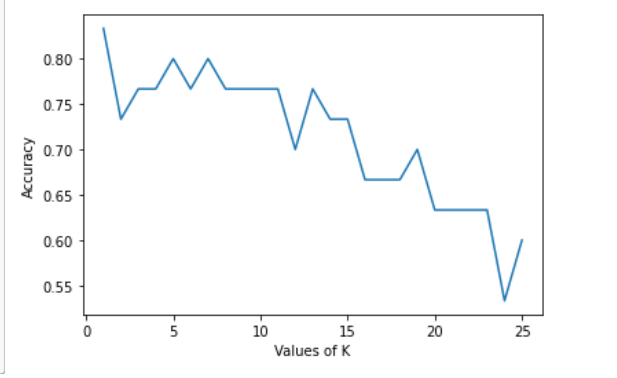
k-nearest neighbors (kNN) is a non-parametric method used in classification. The input consists of the k closest training examples in the feature space. The output is a class membership.

# Table of Observation

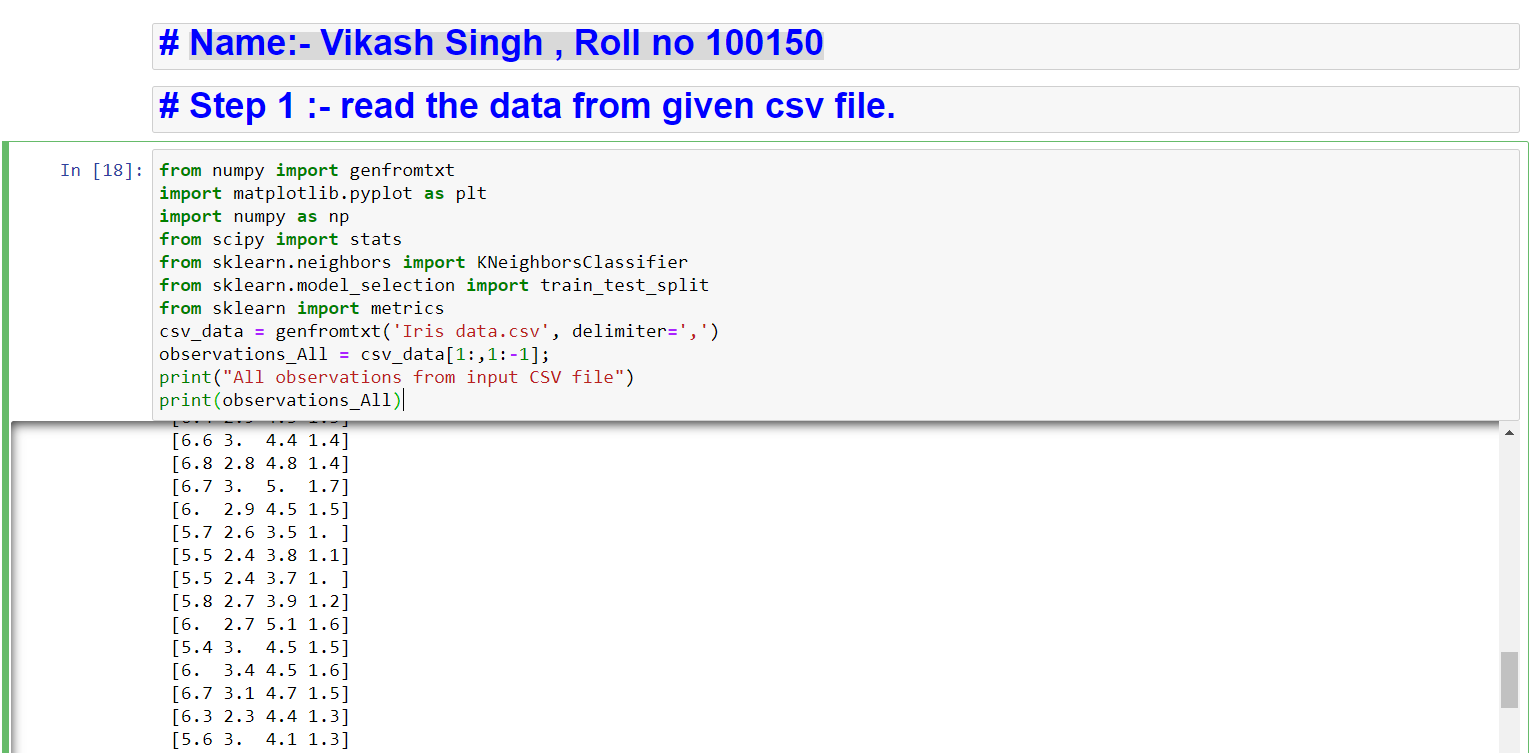
Split the datasets in such as way that for one class of 50 , top 40 will be used for trsaining and rest 10 are used for testing.

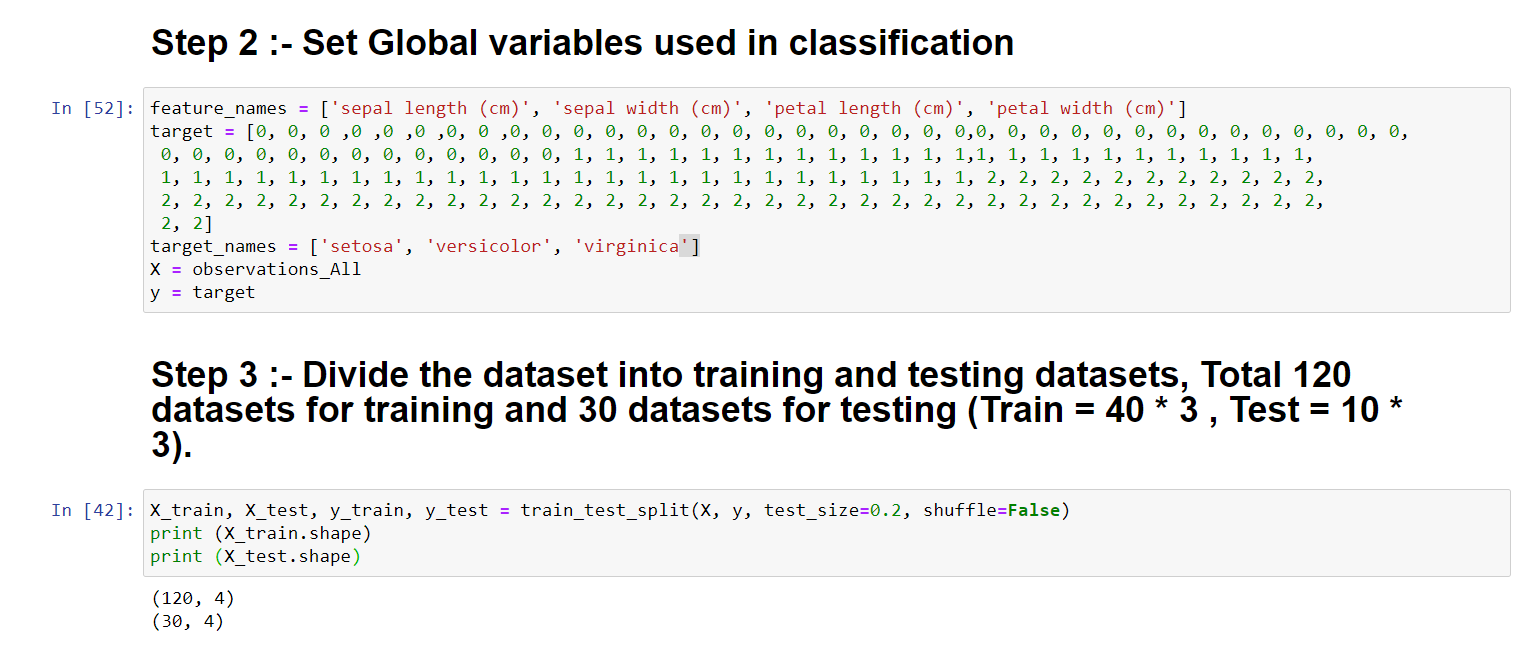
**Observation**: - Accuracy of the model will change if we do normalization and randomization

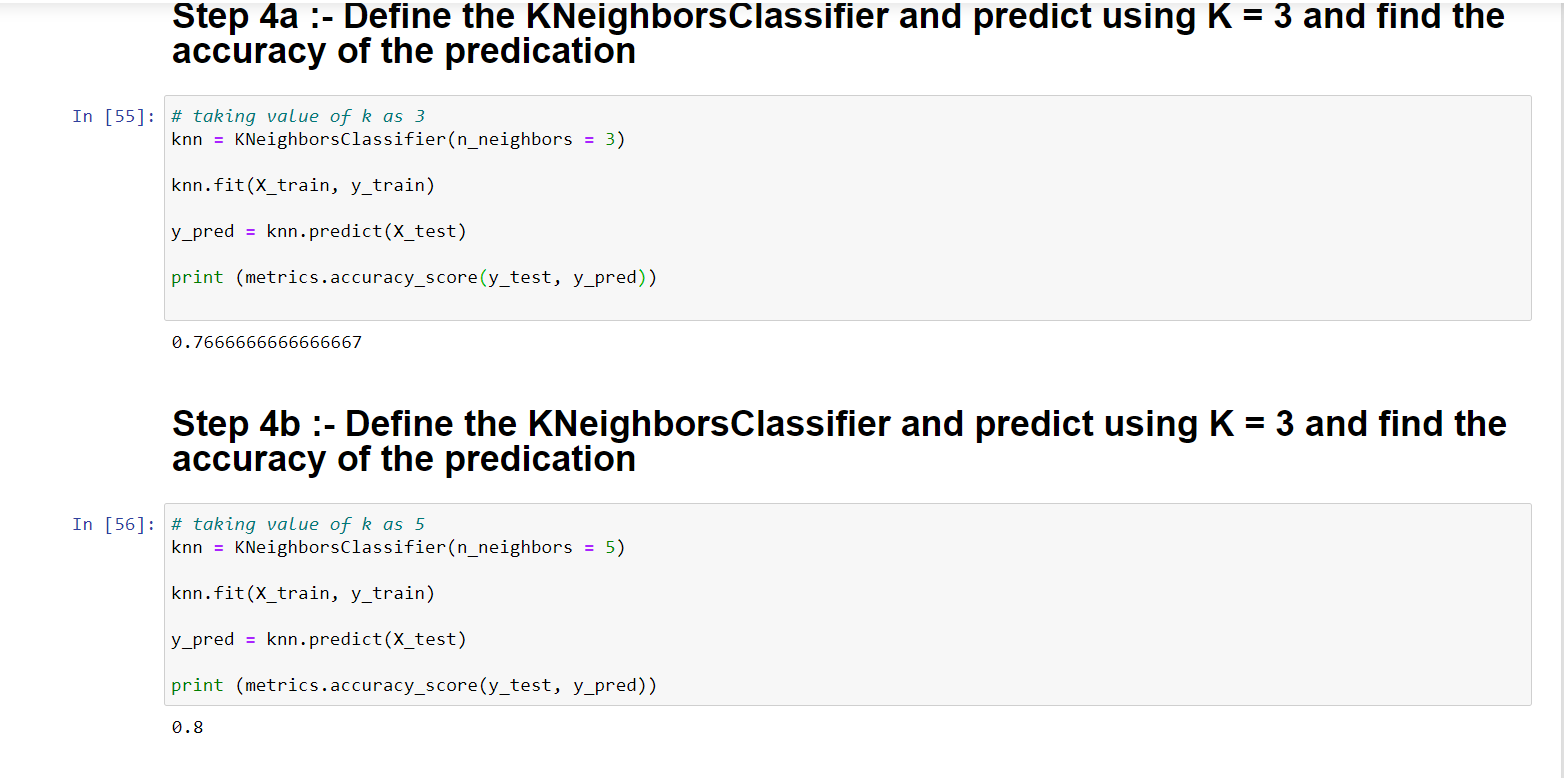
|  |  |
| --- | --- |
| Value of K | Accuracy |
| 3 | 0.76 |
| 5 | 0.8 |
| 7 | 0.8 |



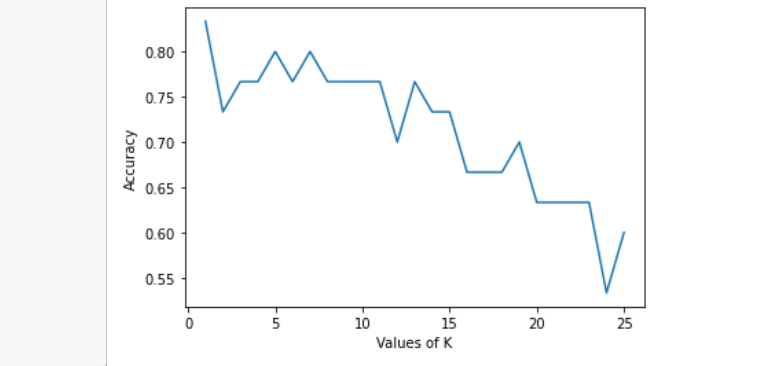
# Python Code











# Scatterplot for K = 12

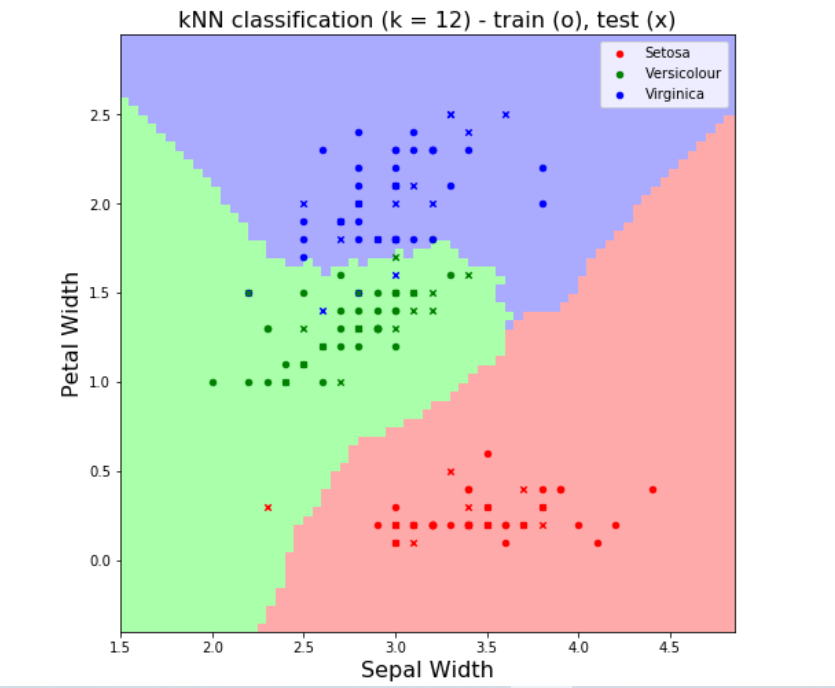


Figure 1:- the scatterplot generated by KNN for K = 12