**KNN Analysis for simulated dataset**

**Creating an artificial dataset:**

For creating an artificial dataset, we used the make\_blobs from the sk.learn datasets and we have also loaded the required packages. Make\_blobs function creates a cluster of points with specific centers.

Text

Description automatically generated

**Splitting the dataset**:

Splitting of dataset into training and testing, by using the train\_test\_split function from the sklearn package I have divided the data into two sets. Which is 80% used for the training and the remaining 20% used for the testing.

Text

Description automatically generated

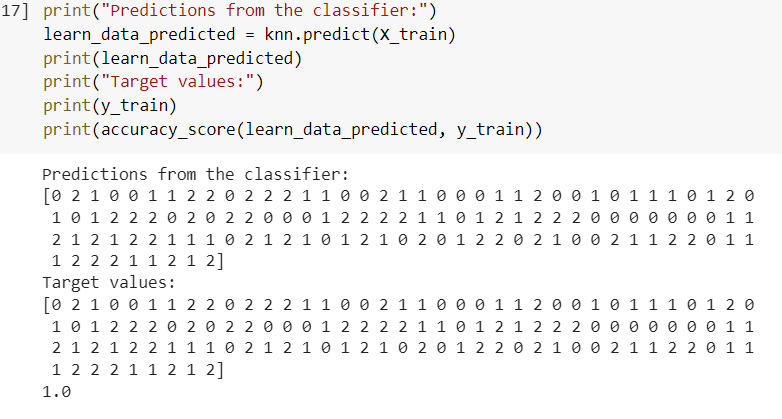
**Performing a KNN analysis of the simulated data:**

Here to perform the KNN analysis by importing the Kneighbhorsclassifier from sklearn package. Using of Kneighorsclassifier determines the how many nearest k neighbhors to look at.

Graphical user interface, text, application

Description automatically generated

Then did some interesting metrics to print like the accuracy, the predictions from the classifier and the target values.



From the above we can say that the predictions from the classifier and the target values are same and the accuracy is 1. It tells us that the model perfectly fits the model, and it will work well on unseen data.

**Evaluate K-nearest neighbors classifier for different values of k and determine the accuracy:**

It is important to see how the classifier works for different values of k and determine the accuracy for different values of k. The values for k from 1-20

Text, application

Description automatically generated

From the output we can say that dataset was well separated. Generally the accuracy of K = 1 suggests that labels were assigned properly but in this case of our dataset is too small, hence the accuracy of k= 1 suggests that the dataset is well separated.

**Plot your different results**:

Here I have created three separate scatterplots which are the following.

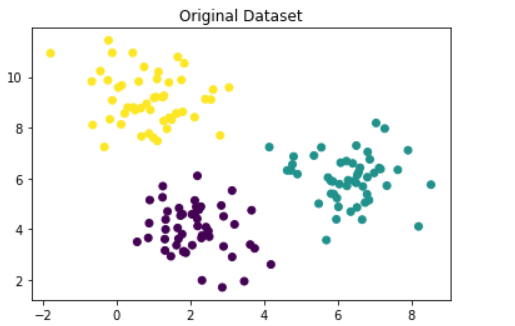
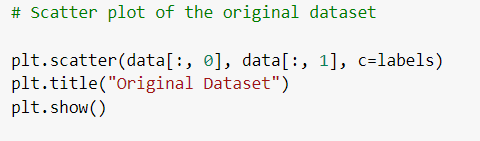
1.Shows the original dataset.

2.Shows the predicted labels on training data.

3.Shows the predicted labels on testing data.

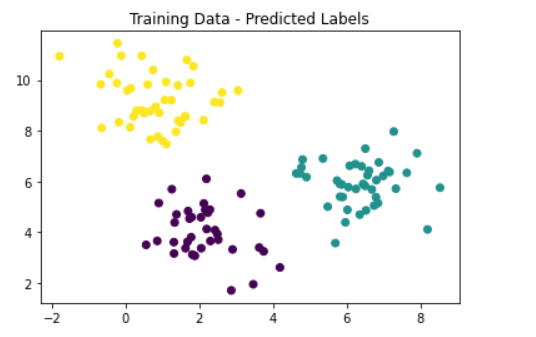
4.Shows the K-nearest neighbhors classifier accuracy using different values of k

**1.Scatter plot of original dataset**



**2.scatter plot of predicted labels on training data**

Text

Description automatically generated

**3.scatter plot of predicted labels on testing**

Text

Description automatically generatedChart, scatter chart

Description automatically generated

**4.Plot showing the K-nearest neighbhors classifier accuracy using different values of k**

A picture containing graphical user interface

Description automatically generated

As discussed above, the accuracy of k is same for the different values of k which tells us that data has been well separated.