**DATE:**

**TASK3: K-Means**

**AIM:** Program to implement k-means algorithm for clustering.

**Data set:** Iris

**Source:** UCI Repository

**I .Clustering using k-means:**

**Description**:

Clustering is a group of objects that belong to same class. In other words similar objects are grouped together and dissimilar objects are grouped in different groups.

**Clustering methods:**

* Partitioning method
* Hierarchical method
* Density-Based method
* Grid – Based method
* Model-Based method
* Constraint – Based method

**1.K-means clustering:**

It is most used algorithm for clustering .It uses Euclidian Distance between any two points to match its similarity.

Euclidian Distance =sqrt ( pow(x1-x2,2) + pow(y1-y2,2) )

**Algorithm: k-means**

The k-means algorithm for partitioning where each clusters center is represented by the mean value of the objects in the cluster.

**Input:**

K: the number of clusters

D: a data set containing n objects

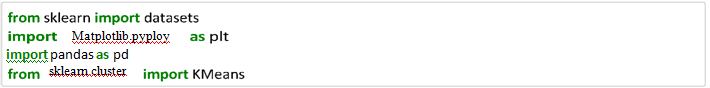
**Output:** A set of n clusters

**Method:**

1. Arbitrarily choose k objects from D as the initial cluster centers.
2. Repeat
3. (re)assign each object to the cluster to which the object is the most similar,based on the mean value of the objects in the cluster;
4. Update the cluster means, i.e; calculate the mean value of the objects for each cluster.;
5. Until no change;

**PROGRAM:**

In [1]:



In [2]:

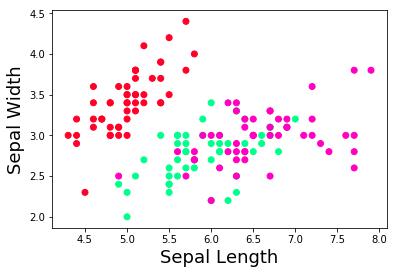


In [5]:



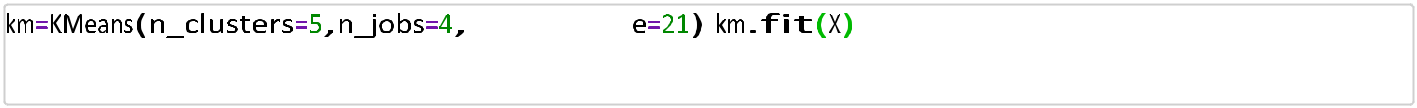
In [6]:

Out[6]:

Text(0, 0.5, 'Sepal Width')



In [7]:



Out[7]:

KMeans(algorithm='auto', copy\_x=True, init='k-means++', max\_iter=300, n\_clusters=5, n\_init=10,

n\_jobs=4, precompute\_distances='auto',

random\_state=21, tol=0.0001, verbose=0)

In [8]:



[[7.43846154 3.13076923]

[5.83953488 2.70930233]

[4.76666667 2.89166667]

[6.53421053 3.04210526]

[5.1875 3.6375]]

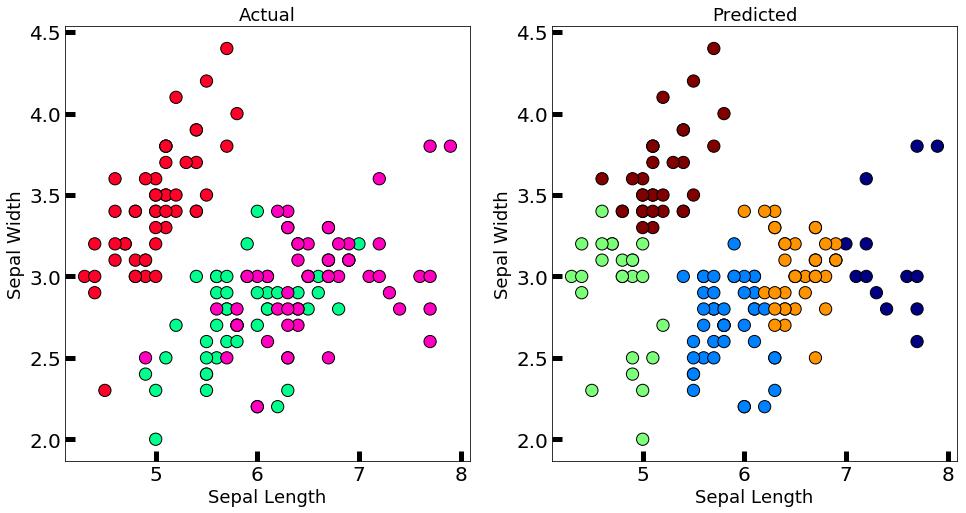
In [10]:



Out[10]:

Text(0.5, 1.0, 'Predic

ted')



**Result:** k-means algorithm implemented successfully.