|  |  |
| --- | --- |
| **Name:** | **Soham Ajit Dahanukar** |
| **Roll No:** | **13** |
| **Class/Sem:** | TE/V |
| **Experiment No.:** | 8 |
| **Title:** | Implementation of any one clustering algorithm using languages like JAVA/ python. |
| **Date of Performance:** |  |
| **Date of Submission:** |  |
| **Marks:** |  |
| **Sign of Faculty:** |  |

**Aim:** To Study and Implement K‐Means algorithm

**Objective:-** Understand the working of K‐Means algorithm and it’s implemention using python.

**Theory:**

In statistics and machine learning, k‐means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Input

K:-number of clusters

D:- data set containing n objects

Output

A set of k clusters

Given k , the k-means algorithm is implemented in 5 steps:

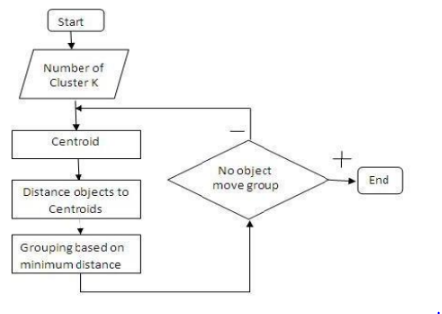
Step 1: Arbitrarily choose k objects from D as the initial cluster centers.

Step 2: Find the distance from each and every object in the dataset with respect to cluster centers

Step 3: Assign each object to the cluster with the nearest seed point based on the mean value of the objects in the cluster.

Step 4: Update the cluster means i,e calculate the mean value of the objects for each cluster.

Step 5: Repeat the procedure, until there is no change in meaning.



Example: d= {2,4,10,12,3,20,30,11,25} k =2

1. Randomly assign mean m1=3 and m2 = 4

Therefore, k1 = {2,3} Therefore, k1 = {4,10,12,20,30,11,25}

2. Randomly assign mean m1=2.5 and m2 = 16

Therefore, k1 = {2,3,4} Therefore, k1 =

{4,10,12,20,30,11,25}

3. Randomly assign mean m1=3 and m2 = 18

Therefore, k1 = {2,3,4,10} Therefore, k1 = {12,20,30,11,25}

4. Randomly assign mean m1=7 and m2 = 25

Therefore, k1 = {2,3,4,10,11,12} Therefore, k1 =

{20,30,25}

5. Randomly assign mean m1=7 and m2 = 25

Therefore, we stop as we are getting same mean

values.

6. Therefore, Final clusters are : k1 = {2,3,4,10,11,12} Therefore, k1 = {20,30,25}

**CODE:**

import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

from sklearn.datasets import make\_blobs

# Generate synthetic data

data, \_ = make\_blobs(n\_samples=300, centers=4, random\_state=42)

# Apply K-Means algorithm

kmeans = KMeans(n\_clusters=4, random\_state=42)

kmeans.fit(data)

# Get cluster centers and labels

centers = kmeans.cluster\_centers\_

labels = kmeans.labels\_

# Visualize the data and clusters

plt.scatter(data[:, 0], data[:, 1], c=labels, cmap='viridis', edgecolors='k')

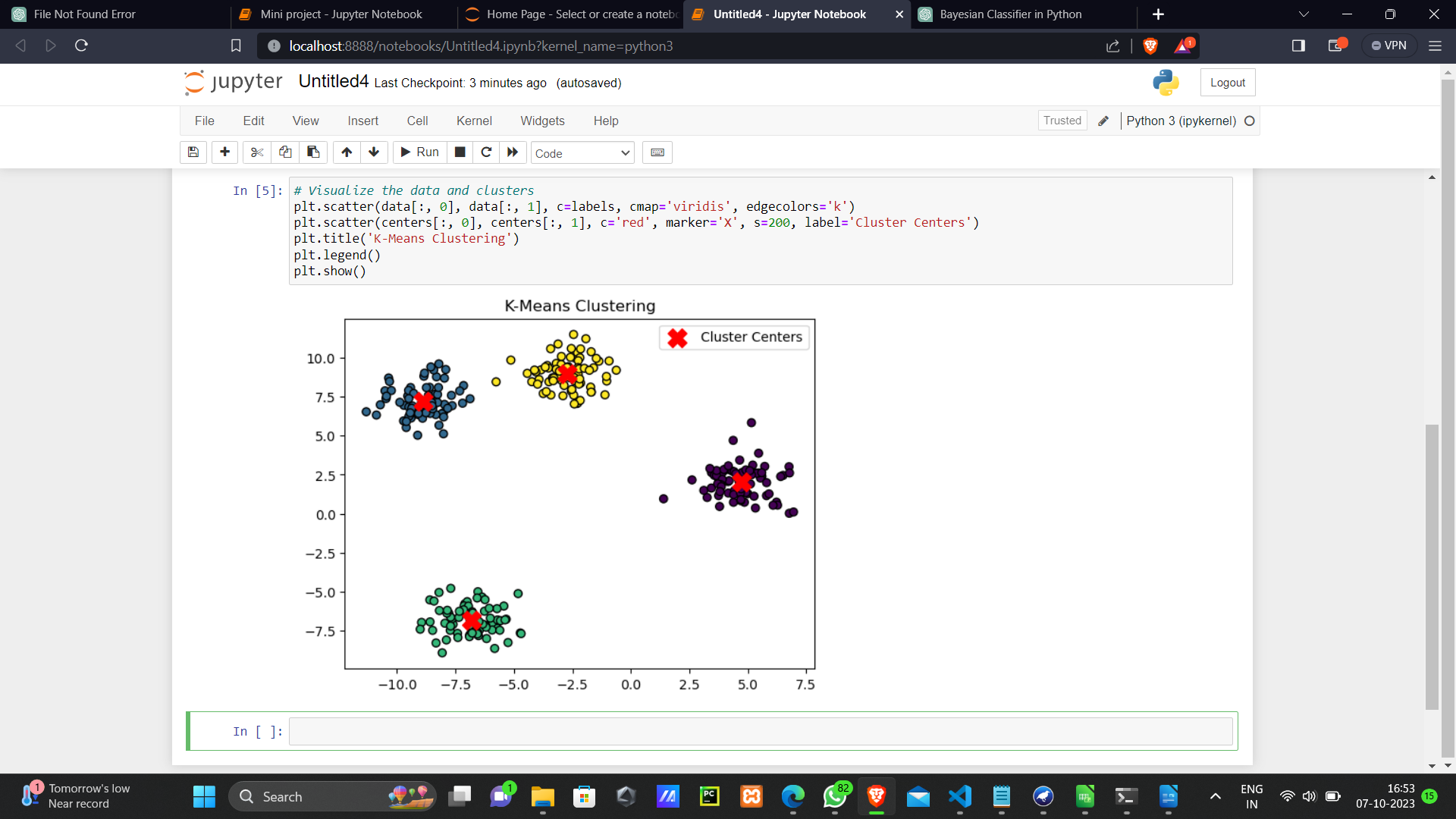
plt.scatter(centers[:, 0], centers[:, 1], c='red', marker='X', s=200, label='Cluster Centers')

plt.title('K-Means Clustering')

plt.legend()

plt.show()

**OUTPUT:**



**CONCLUSION:**