|  |  |
| --- | --- |
| **EX.NO 9 DATA & TEXT CLUSTERING USING K-MEANS ALGORITHM** | |
| IN [1] | import pandas as pd  import numpy as np  import matplotlib.pyplot as plt  import seaborn as sns |
| IN [2] | from sklearn import datasets  wine = datasets.load\_wine() |
| IN [3] | df = pd.DataFrame(wine.data, columns = wine.feature\_names)  df['target']=wine.target  df.to\_csv("wine.csv") |
| IN [4] | print(wine.feature\_names) |
| IN [5] | print(wine.target\_names) |
| IN [6] | print(wine.data[0:5]) |
| IN [7] | print(wine.target) |
| IN [8] | print(wine.data.shape) |
| IN [9] | print(wine.target.shape) |
| IN [10] | from sklearn.model\_selection import train\_test\_split  X\_train, X\_test, y\_train, y\_test = train\_test\_split(wine.data, wine.target, test\_size=0.3) |
| IN [11] | X=df  from sklearn.cluster import KMeans  kmeans = KMeans(n\_clusters=2, random\_state=0)  kmeans.fit(X) |
| IN [12] | from sklearn.neighbors import KNeighborsClassifier  import sklearn.metrics as metrics  from sklearn.metrics import accuracy\_score  Clusters = []  Accuracy = []  for i in range(1,20):  knn = KNeighborsClassifier(n\_neighbors=i)  knn.fit(X\_train, y\_train)  y\_pred = knn.predict(X\_test)  Clusters.append(i)  Accuracy.append(metrics.accuracy\_score(y\_test, y\_pred))  print(Clusters)  print(Accuracy) |
| IN [13] | kmeans = KMeans(n\_clusters=3, random\_state=0)  kmeans.fit(X) |
| IN [14] | from sklearn.preprocessing import LabelEncoder  le = LabelEncoder() |
| IN [15] | import matplotlib.pyplot as plt  plt.title("Clusters (k) Vs Accuracy\_score")  plt.xlabel("Clusters (k)")  plt.xlim(0,20,1)  plt.ylabel("Accuracy\_score")  plt.plot(Clusters, Accuracy)  plt.show() |

**OUTPUT**

|  |  |
| --- | --- |
| OUT[15] |  |