PRACTICAL NUMBER: 6

Aim: Implement the classification model using clustering for following techniques with Kmeans clustering with Prediction, Test Score and Confusion Matrix.

Code:

```
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.cluster import KMeans
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import sklearn.metrics as sm
import pandas as pd
import numpy as np

#Loading the iris dataset
iris = datasets.load_iris()
X = pd.DataFrame(iris.data)
X.columns = ['Sepal_Length','Sepal_Width','Petal_Length','Petal_Width']
y = pd.DataFrame(iris.target)
y.columns = ['Targets']
X_train,X_test,y_train,y_test = (train_test_split(X,y,test_size=0.3))
#n_clusters: The number of clusters to form as well as the number of centroids to generate.
model = KMeans(n_clusters=3)
model.fit(X)
y_pred = model.predict(X_test)

print("Model labels:")
print(model.labels_)
print("Accuracy is: ")
print(classification_report(y_test,y_pred))
print("Confusion matrix is:")
print(confusion_matrix(y_test,y_pred))

plt.figure(figsize=(14,7))
colormap = np.array(['red','cyan','black'])
plt.subplot(1,2,1)
plt.scatter(X.Petal_Length,X.Petal_Width, c=colormap[y.Targets], s = 40)
plt.title("Real clusters")
plt.subplot(1,2,2)
plt.scatter(X.Petal_Length,X.Petal_Width, c=colormap[model.labels_], s= 40)
plt.title("K Means Cluster")
plt.show()
```

Output:

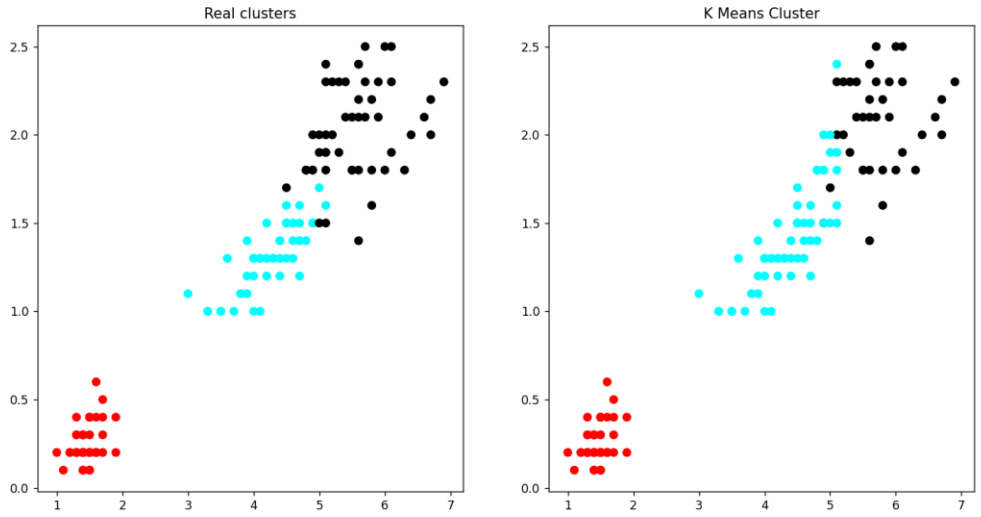
```
*Python 3.7.9 Shell*
File  Edit  Shell  Debug  Options  Window  Help
Python 3.7.9 (tags/v3.7.9:13c94747c7, Aug 17 2020, 18:58:18) [MSC v.1900 64 bit
(AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: D:/MSc IT/Sem3/ML Pracs/Prac6.py ==========
Model labels:
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2 2 2 2
 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2
 2 1]
Accuracy is:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        11
           1       0.81      1.00      0.90        22
           2       1.00      0.58      0.74        12

    accuracy                           0.89        45
   macro avg       0.94      0.86      0.88        45
weighted avg       0.91      0.89      0.88        45

Confusion matrix is:
[[11  0  0]
 [ 0 22  0]
 [ 0  5  7]]
```

Figure 1



1. Explain Distance method (Euclidean method)

2. Explain K means Clustering Algorithm