8. Apply EM algorithm to cluster a set of data stored in a .csv file. Use the same data set for clustering using K-means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes / API in the program.

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

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']

model = KMeans(n_clusters=3)
model.fit(x)
model.labels_
plt.figure(figsize=(14,7))
Colormap = np.array(['red', 'lime', 'black'])

plt.subplot(1,2,1)
plt.scatter(X.Petal_Length, X.Petal_Width, c=colormap[y.Targets],
                                                    s=40)
```

Teacher's Signature : _____

```
plt.title('Real Classification')

plt.subplot(1,2,2)
plt.scatter(X.petal_length, X.Petal_width, c=colormap[model.labels_],
                                                    s=40)

plt.title('K Mean Classification')
plt.figure(figsize=(14,7))

predy = np.choose(model.labels_, [0,1,2]).astype(np.int64)
print(predy)

plt.subplot(1,2,1)
plt.scatter(X.Petal_Length, X.Petal_width, c=colormap[y.Targets],
                                                    s=40)

plt.title('Real classification')
plt.subplot(1,2,2)
plt.scatter(X.Petal_length, X.Petal_width, c=colormap[predy], s=40)
plt.title('K Mean Classification')

print('The accuracy score of K-Mean: ', sm.accuracy_score(y,
                                                    model.labels_))
print('The confusion matrix of K-Mean:', sm.confusion_matrix(y,
                                                    model.labels_))

from sklearn import preprocessing
scaler = preprocessing.StandardScaler()
scaler.fit(X)
Xsa = scaler.transform(X)
XS = pd.DataFrame(Xsa, column = X.columns)
```

Teacher's Signature : _____

```
from sklearn.mixture import GaussianMixture
gmm = GaussianMixture(n_components = 3)
gmm.fit(xs)

y_cluster_gmm = gmm.predict(xs)

plt.subplot(2,2,3)
plt.scatter( x.Petal_length, x.Petal_width, c=colormap[y_cluster_gmm],
                                                    s=40)
plt.title('GMM classification')

print('The accuracy score of EM:', sm.accuracy_score(y, y_cluster_
                                                    gmm))
print('The Confusion matrix of EM:', sm.confusion_matrix(y,
                                                    y_cluster_gmm))
```

Abigal Sharal Mathias, 4MT17CS005

Output:

[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 2 2 1 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 1 2 1 1 1 1 2 1 1 1 1
1 1 2 2 1 1 1 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1
2 1 1 1 2 1 1 1 2 1 1 2]

The accuracy score of K-Mean: 0.44

The confusion matrix of K-Mean: [[50 0 0]
[0 2 48]
[0 36 14]]

The accuracy score of EM: 0.3333333333333333

The confusion matrix of EM: [[0 50 0]
[45 0 5]
[0 0 50]]