“In The Name Of God”

HW#5 Report

Professor: Dr.Azmifar

Presenters:

Narges Dehghan

Masome Jafari

Table of Contents

Abstract 1

Kmeans 3

GMM 21

PCA 32

Conclusion 34

Abstract

This assignment encompasses a comprehensive exploration of implement Kmeans and GMM and how to visualize the resulted probability density functions., fundamental techniques in the realm of machine learning and data analysis. The primary objectives are twofold: to understand the fundamental concepts and application of these methods and to apply this knowledge to real-world datasets.

K-means

K-means is a popular machine learning and data mining algorithm that discovers potential clusters within a dataset. Finding these clusters in a dataset can often reveal interesting and meaningful structures underlying the distribution of data. K-means clustering has been applied to many problems in science and still remains popular today for its simplicity and effectiveness.

aims to evaluate the clustering accuracy of the K-means algorithm on six datasets. The code is organized into a class named.

Our code utilizes a custom implementation of the K-Means algorithm encapsulated in the **YourKMeansClass**. The following key components are implemented:

1. **Random Initialization of Centroids:** Initial cluster centroids are randomly selected.
2. **Calculation of Distances:** Euclidean distances are used to find the closest centroid for each data point.
3. **Update of Centroids:** Centroids are updated based on the mean of the points in each cluster.
4. **Elbow Method Plotting:** An elbow plot is generated to help determine an optimal number of clusters (K) based on the within-cluster sum of squares (WCSS).

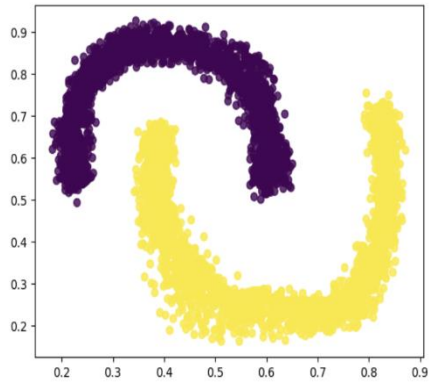
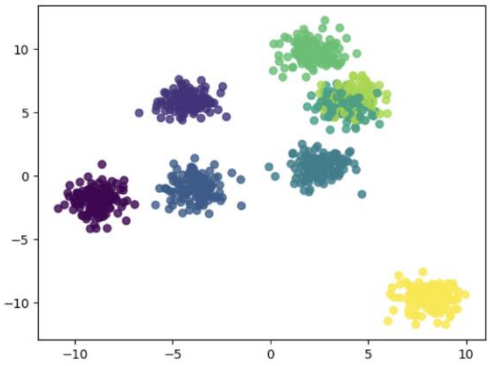
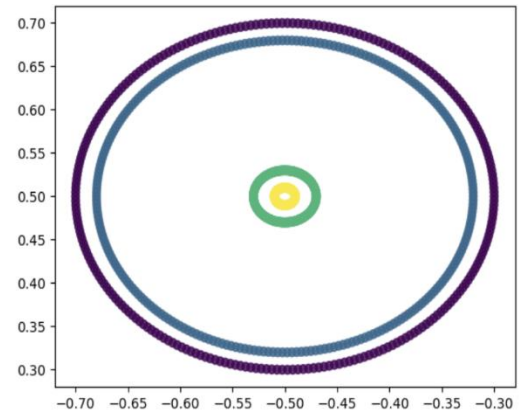
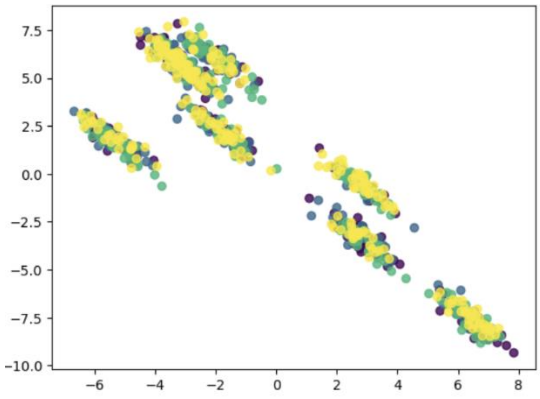
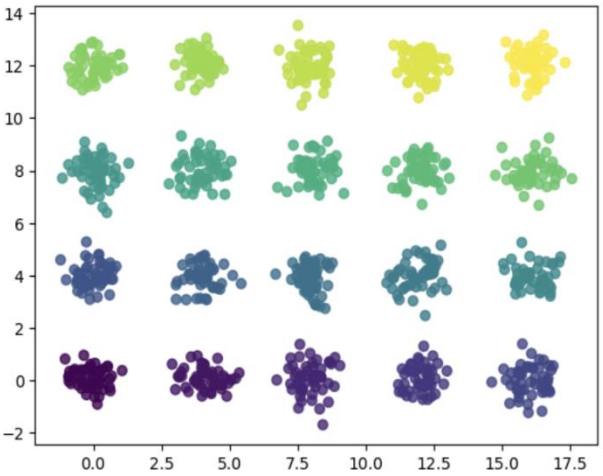
Elbow Method: The elbow method is employed to visually inspect the trade-off between the number of clusters and WCSS. The point where the WCSS begins to show diminishing returns helps identify a suitable K for clustering.

Clustering Accuracy: Adjusted Rand Index (ARI) is used to measure the accuracy of the clustering. ARI is a metric that assesses the similarity between true labels and predicted labels, accounting for permutations of label assignments.

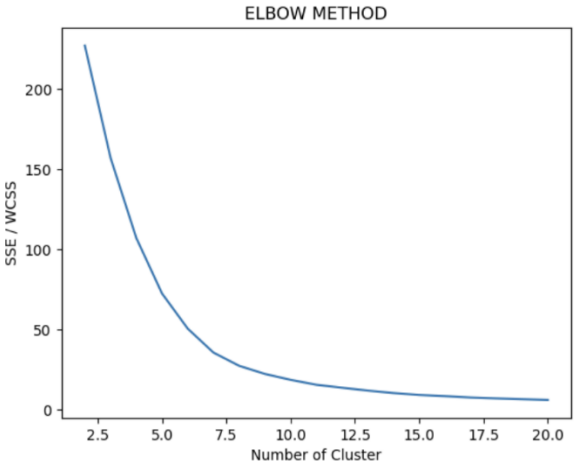
Results: The code iterates over a range of K values (from 2 to 20) and evaluates the clustering accuracy for each K. The Elbow Method plot aids in identifying an optimal K. Clustering accuracy is reported for each K.

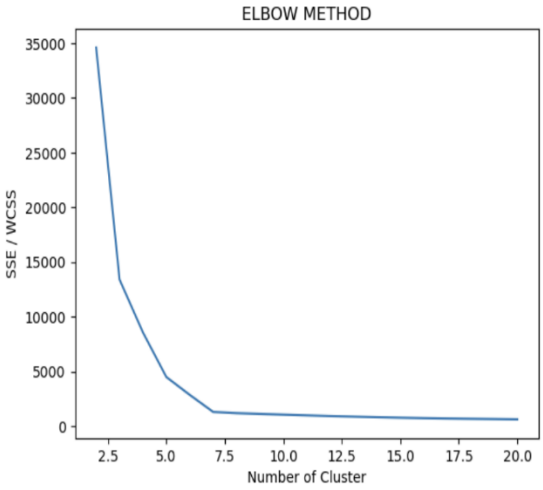
Conclusion: The analysis provides insights into the performance of K-Means clustering on the dataset. The Elbow Method aids in selecting a reasonable number of clusters, and the clustering accuracy metric quantifies the quality of the clustering assignments.

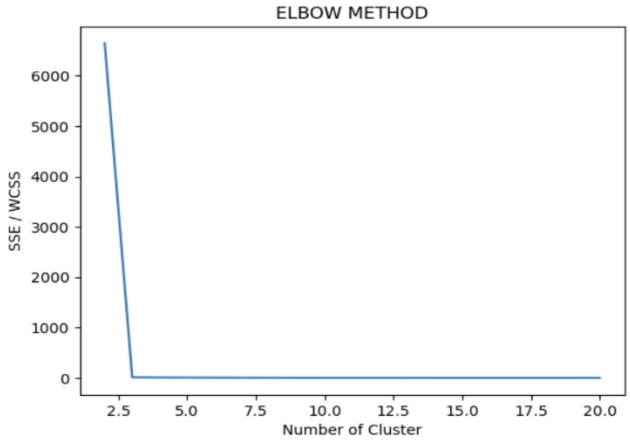
First we load all 6 dataset and plot of them, Perform K-means clustering on all given datasets using euclidean distance, and so on. We report all resolt of this part as the follow.

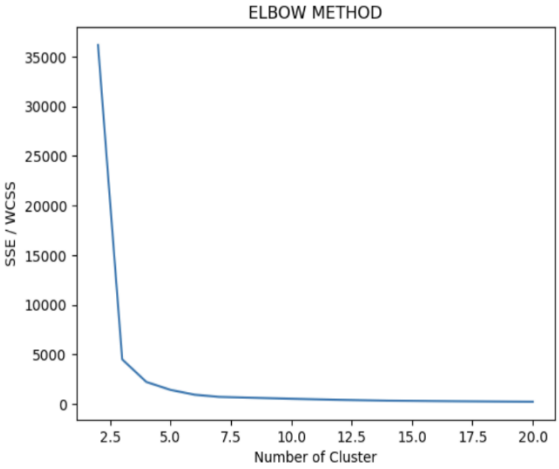






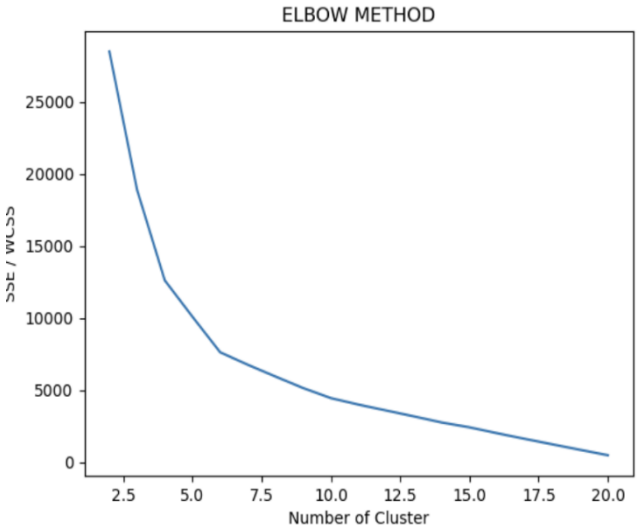
Now we use elbow method and Davies Bouldin Index nd report the plots:

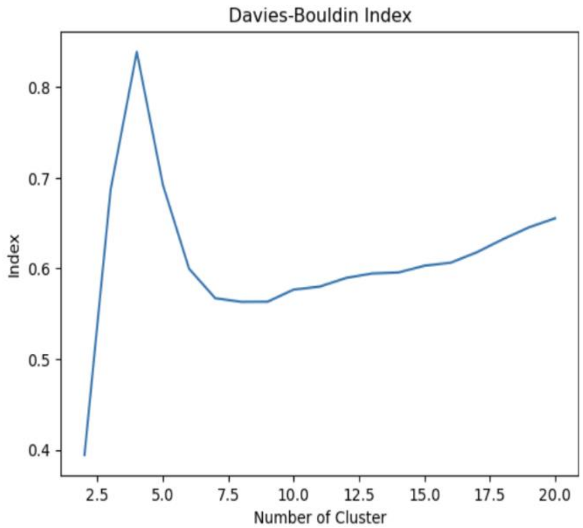


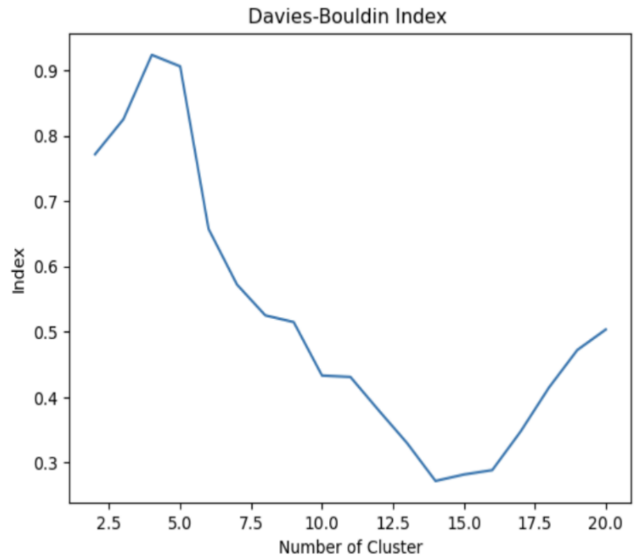


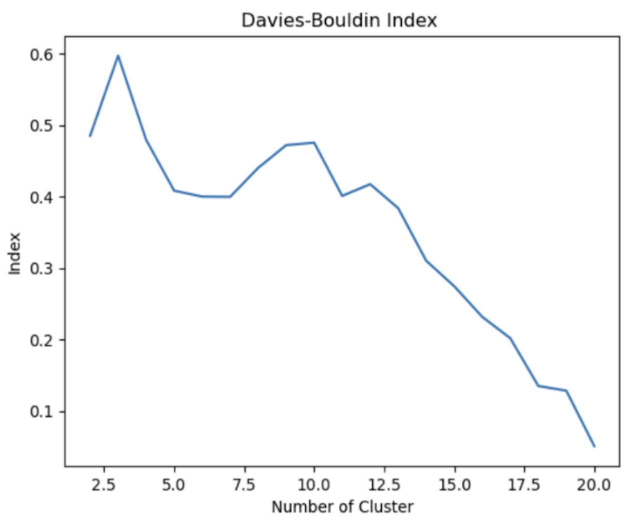


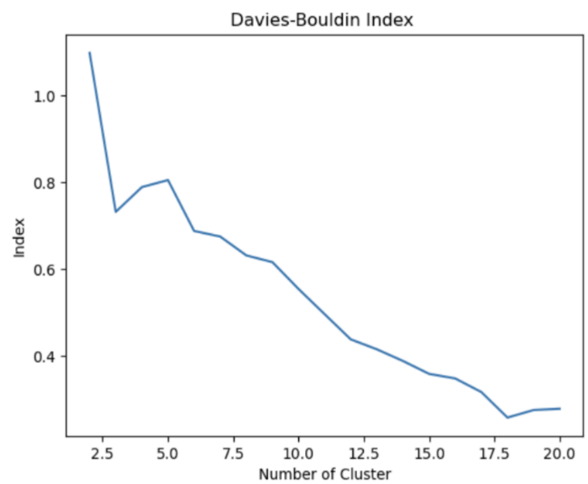


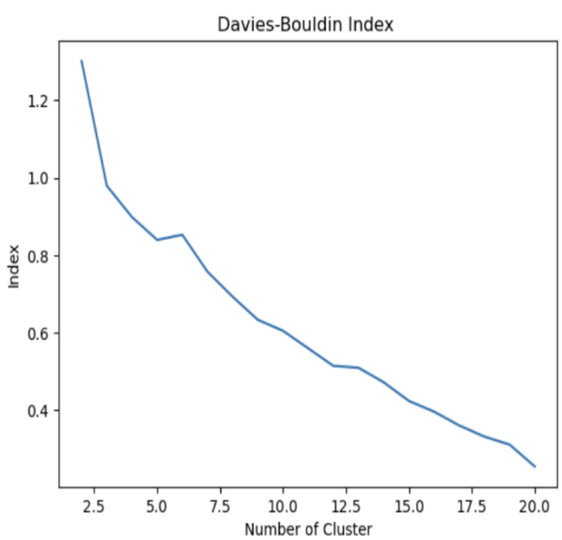


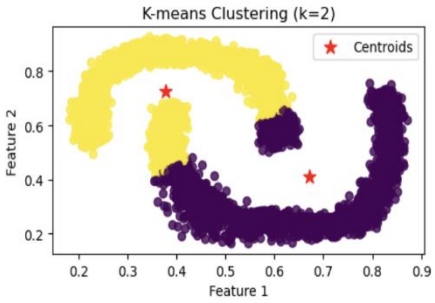

Davies-Bouldin Index
Index
Number of Cluster

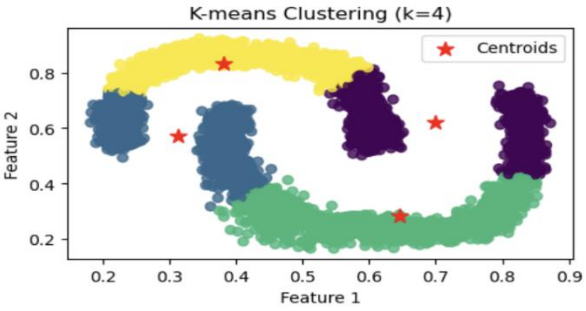

Davies-Bouldin Index
Index
Number of Cluster

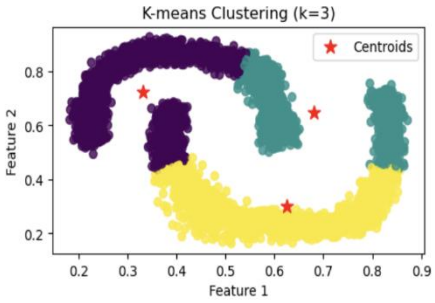

Davies-Bouldin Index
Index
Number of Cluster

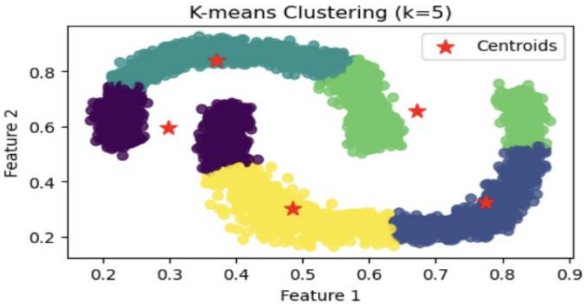

Davies-Bouldin Index
Index
Number of Cluster

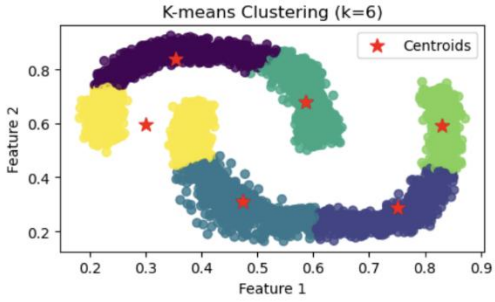

Davies-Bouldin Index
Index
Number of Cluster

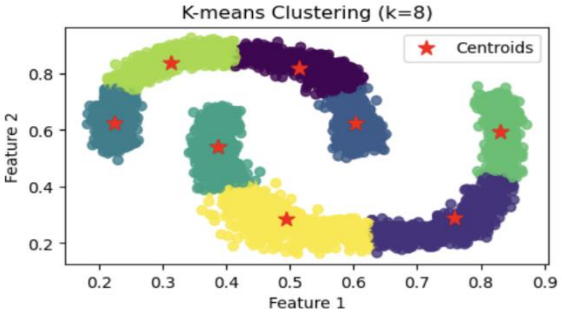

K-means Clustering (k=2)
Centroids
Feature 1
Feature 2

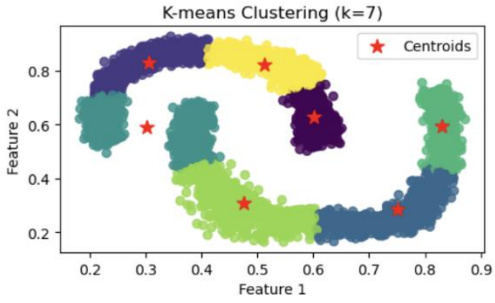

K-means Clustering (k=4)
Centroids
Feature 1
Feature 2

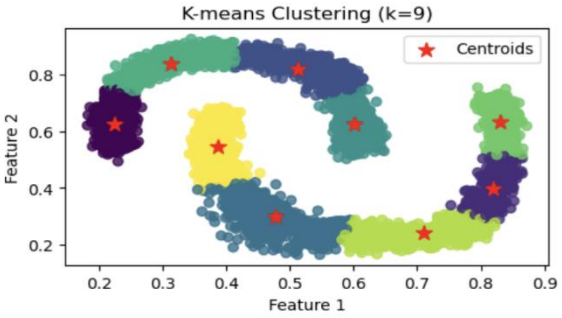

K-means Clustering (k=3)
Centroids
Feature 1
Feature 2

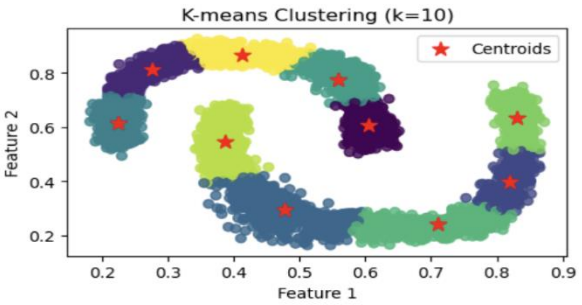

K-means Clustering (k=5)
Centroids
Feature 1
Feature 2

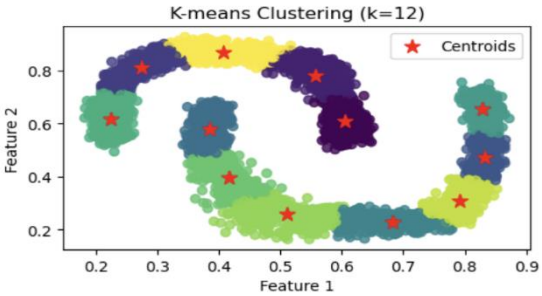

K-means Clustering (k=6)
Centroids
Feature 1
Feature 2

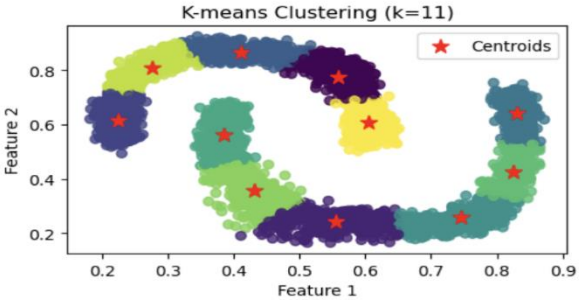

K-means Clustering (k=8)
Centroids
Feature 1
Feature 2

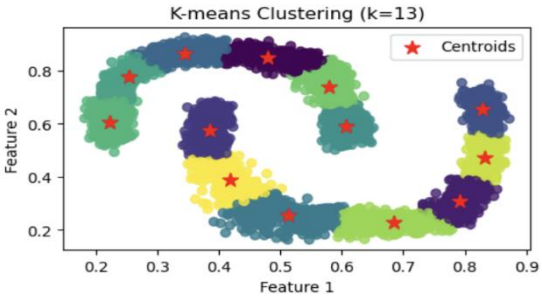

K-means Clustering (k=7)
Centroids
Feature 1
Feature 2

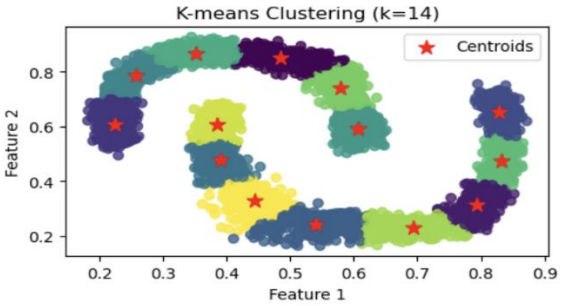

K-means Clustering (k=9)
Centroids
Feature 1
Feature 2

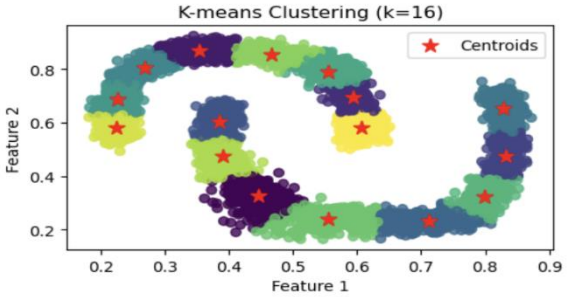

K-means Clustering (k=10)
Centroids
Feature 1
Feature 2

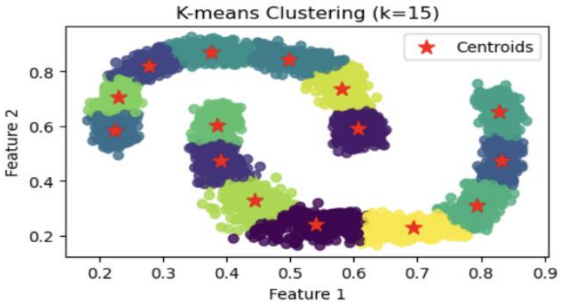

K-means Clustering (k=12)
Centroids
Feature 1
Feature 2

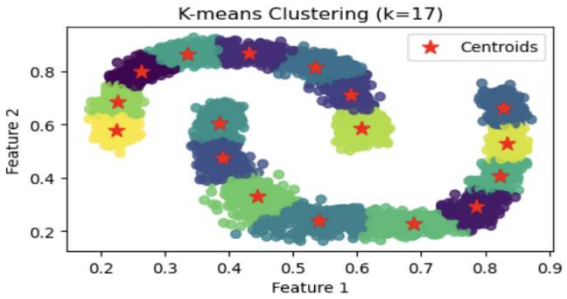

K-means Clustering (k=11)
Centroids
Feature 1
Feature 2

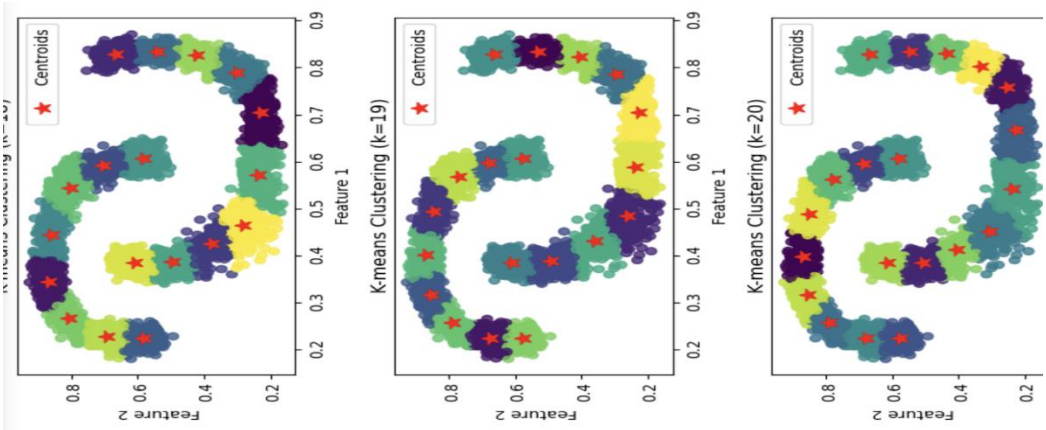

K-means Clustering (k=13)
Centroids
Feature 1
Feature 2

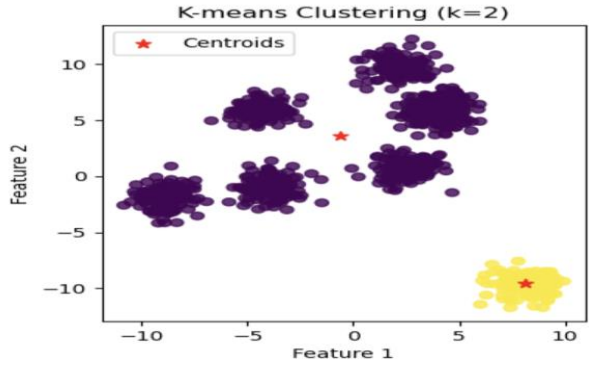

K-means Clustering (k=14)
Centroids
Feature 1
Feature 2

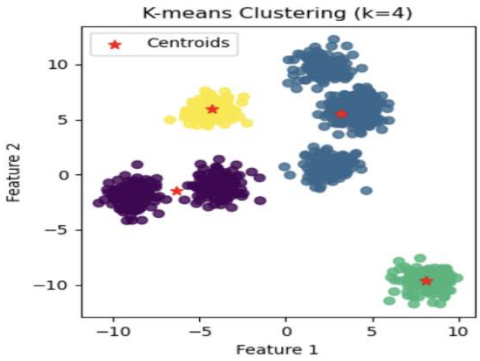

K-means Clustering (k=16)
Centroids
Feature 1
Feature 2

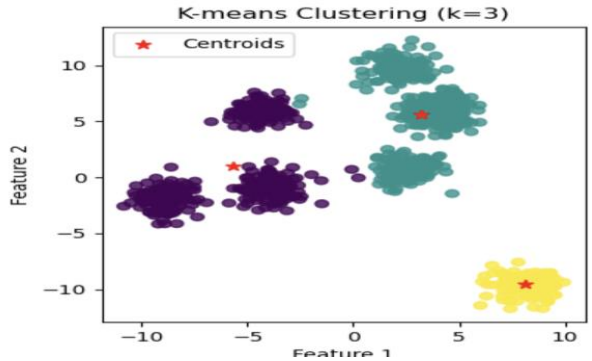

K-means Clustering (k=15)
Centroids
Feature 1
Feature 2

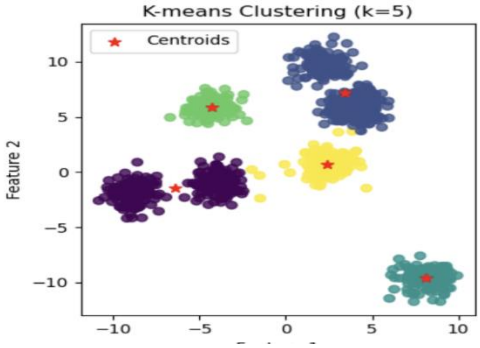

K-means Clustering (k=17)
Centroids
Feature 1
Feature 2

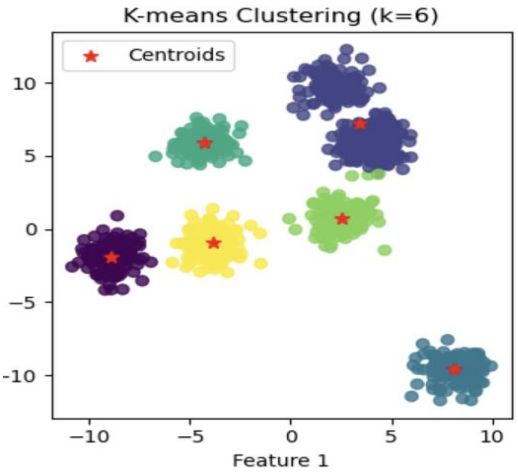

Centroids
Feature 1
Feature 2
K-means Clustering (k=19)
K-means Clustering (k=20)

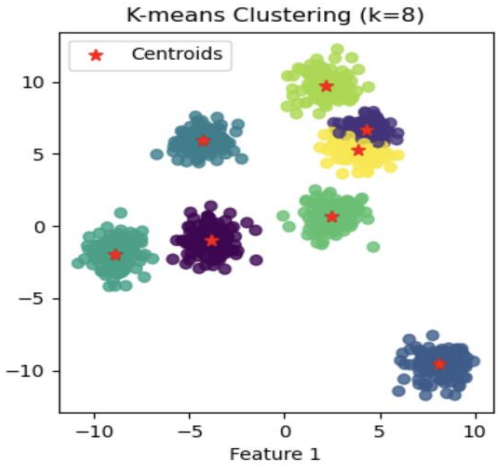

K-means Clustering (k=2)
Centroids
Feature 1
Feature 2

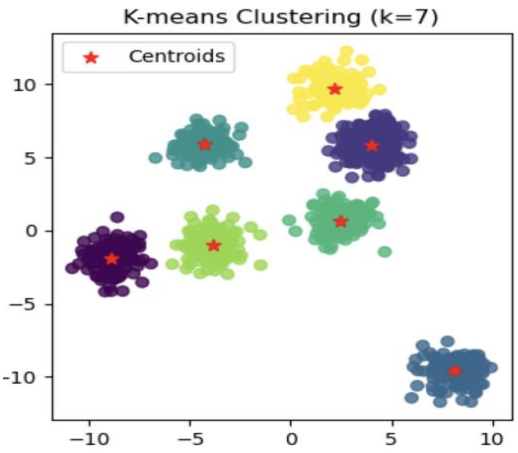

K-means Clustering (k=4)
Centroids
Feature 1
Feature 2

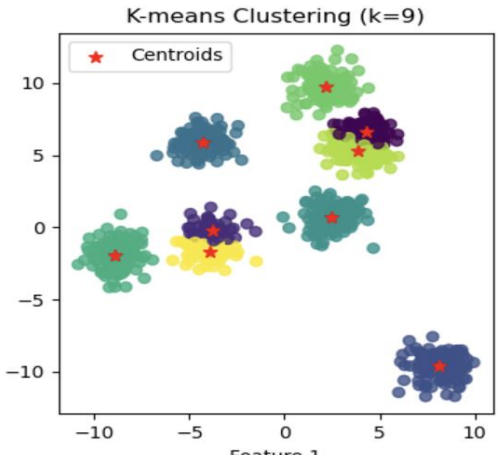

K-means Clustering (k=3)
Centroids
Feature 2

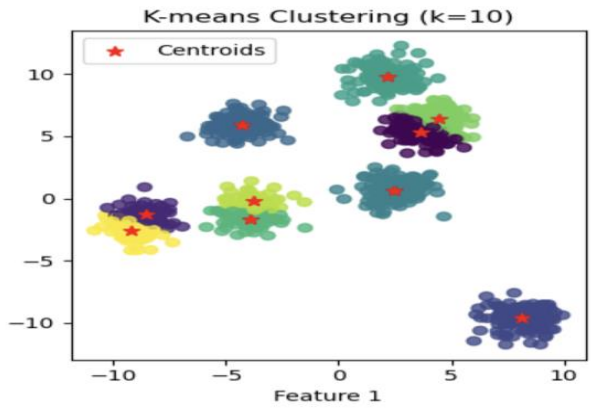

K-means Clustering (k=5)
Centroids
Feature 2

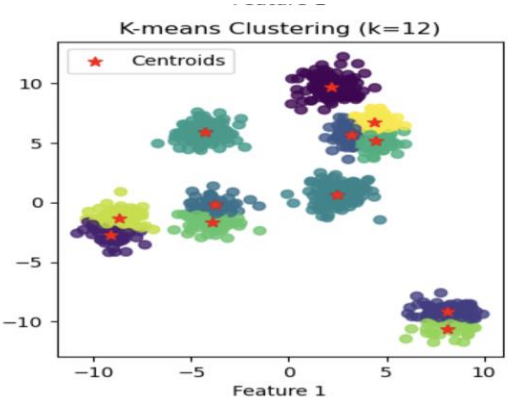

K-means Clustering (k=6)
Centroids
Feature 1

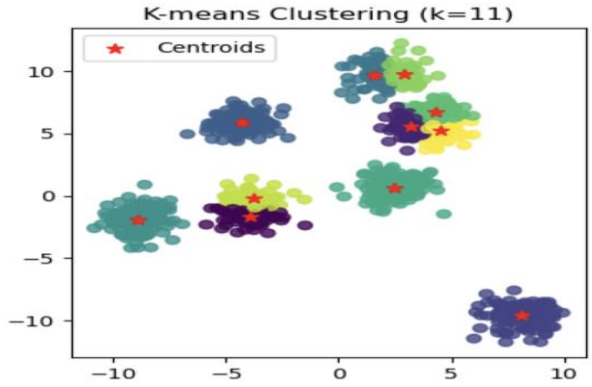

K-means Clustering (k=8)
Centroids
Feature 1

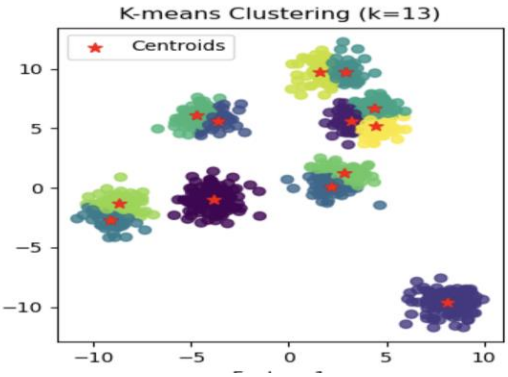

K-means Clustering (k=7)
Centroids

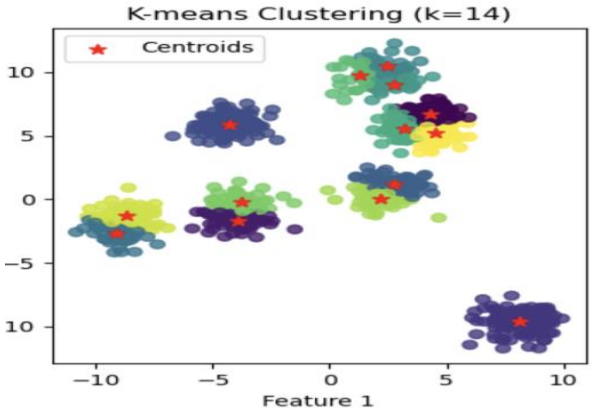

K-means Clustering (k=9)
Centroids

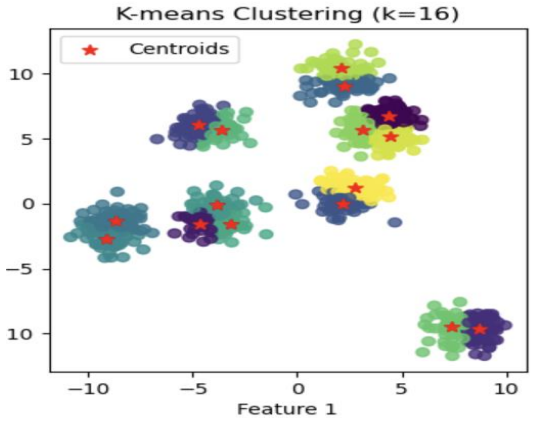

K-means Clustering (k=10)
Centroids
Feature 1

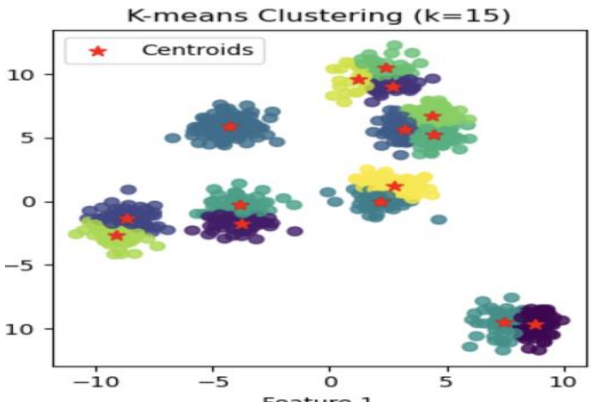

K-means Clustering (k=12)
Centroids
Feature 1

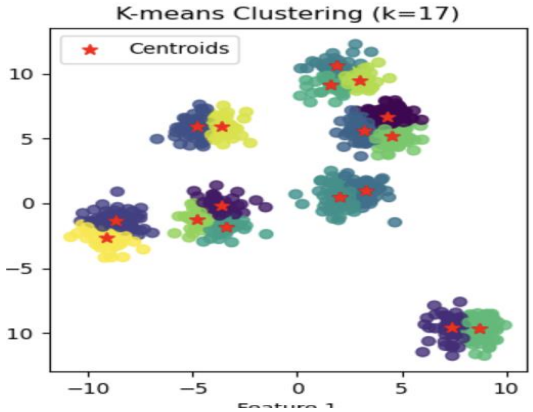

K-means Clustering (k=11)
Centroids

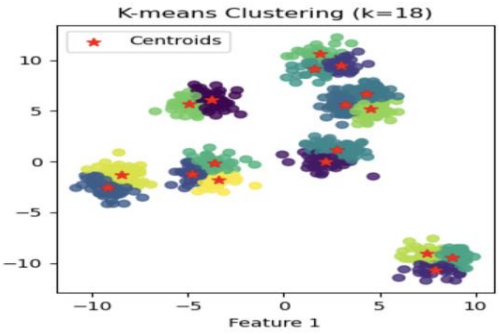

K-means Clustering (k=13)
Centroids

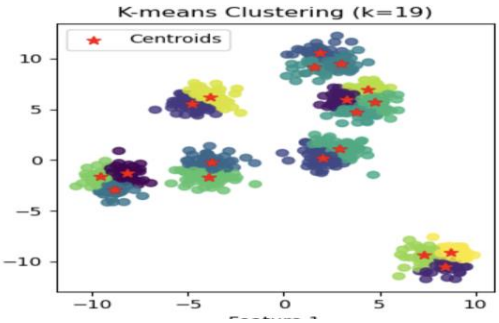

K-means Clustering (k=14)
Centroids
Feature 1

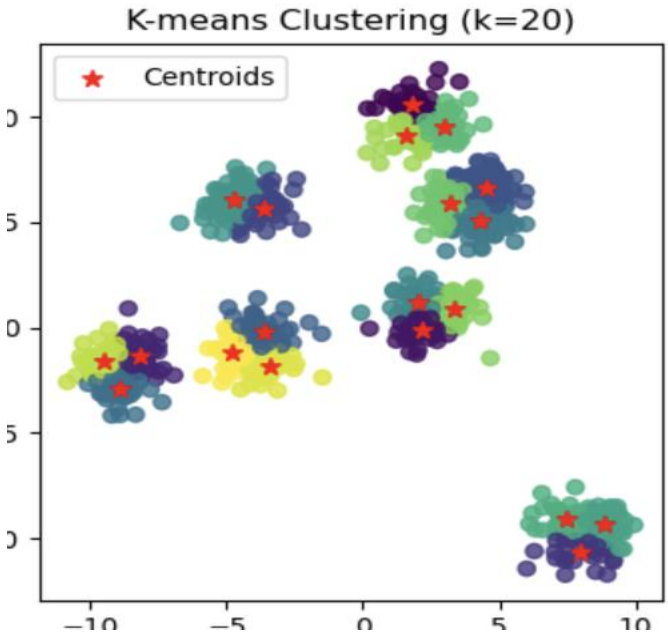

K-means Clustering (k=16)
Centroids
Feature 1

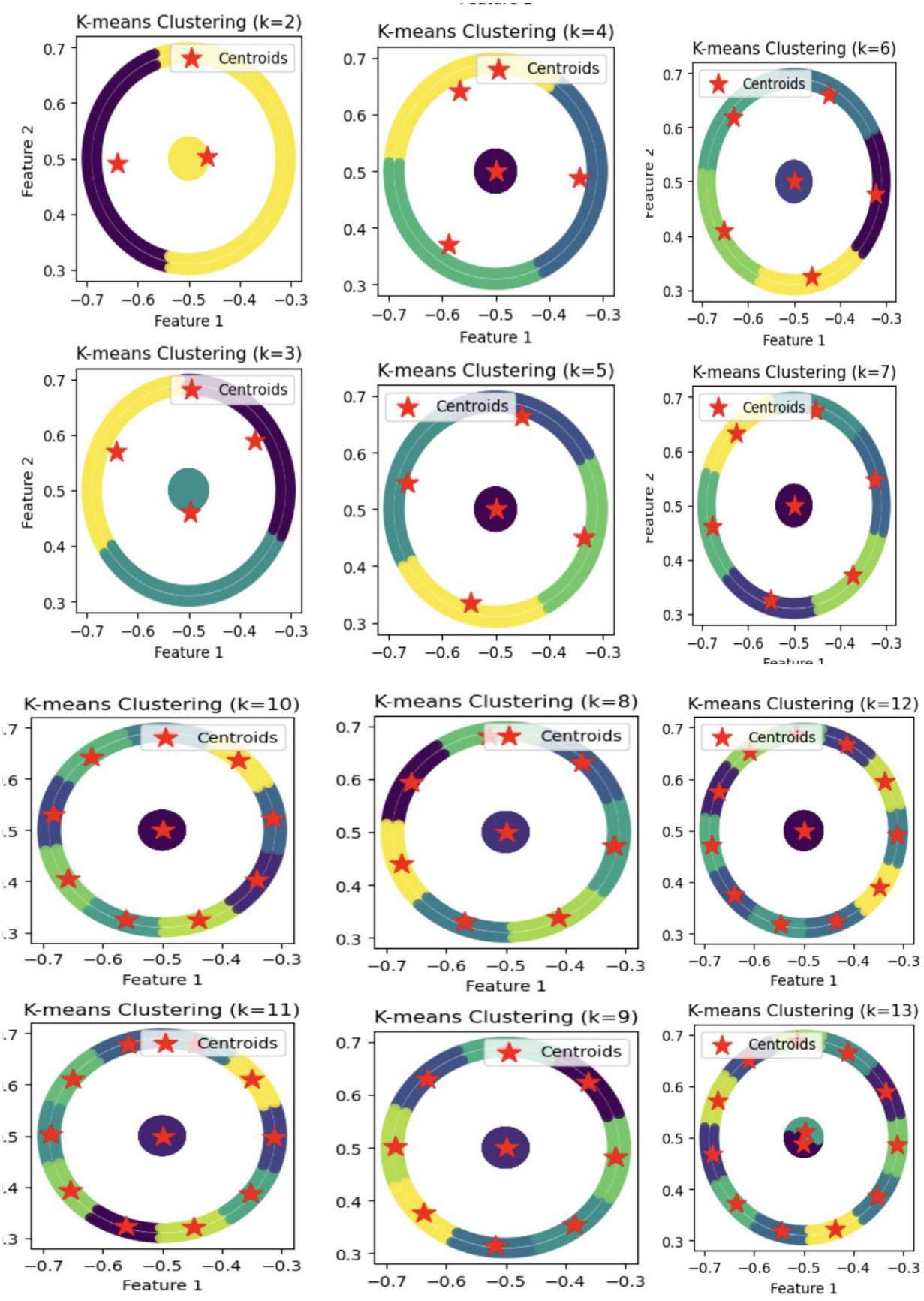

K-means Clustering (k=15)
Centroids

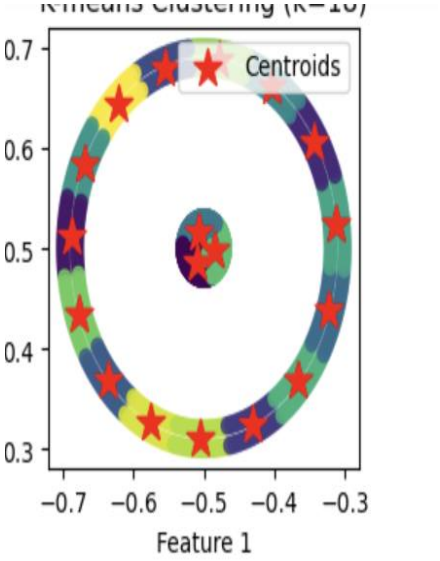

K-means Clustering (k=17)
Centroids

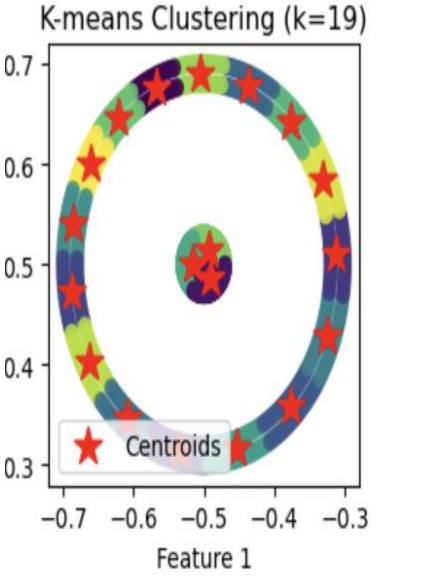

K-means Clustering (k=18)
Centroids
10
5
0
−5
−10
−10
−5
0
5
10
Feature 1

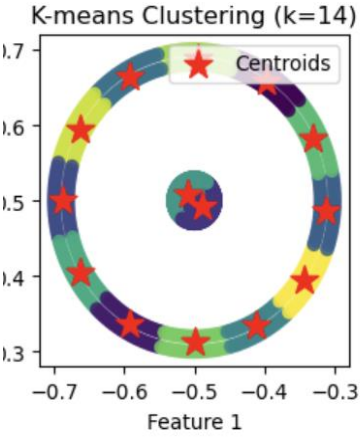

K-means Clustering (k=19)
Centroids
10
5
0
−5
−10
−10
−5
0
5
10

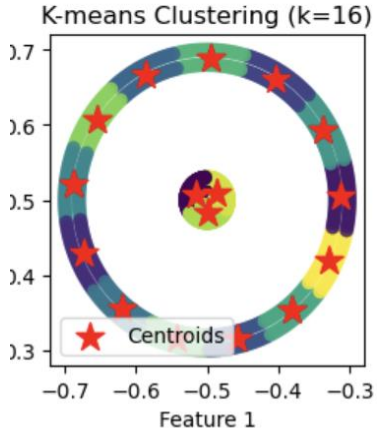

K-means Clustering (k=20)
Centroids
−10
−5
0
5
10

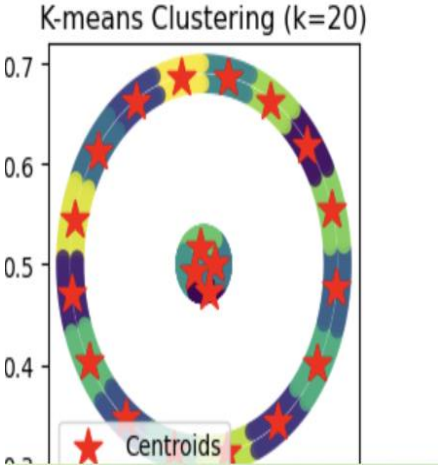

K-means Clustering (k=2)
K-means Clustering (k=4)
K-means Clustering (k=6)
K-means Clustering (k=3)
K-means Clustering (k=5)
K-means Clustering (k=7)
K-means Clustering (k=10)
K-means Clustering (k=8)
K-means Clustering (k=12)
K-means Clustering (k=11)
K-means Clustering (k=9)
K-means Clustering (k=13)
Centroids
Feature 1
Feature 2

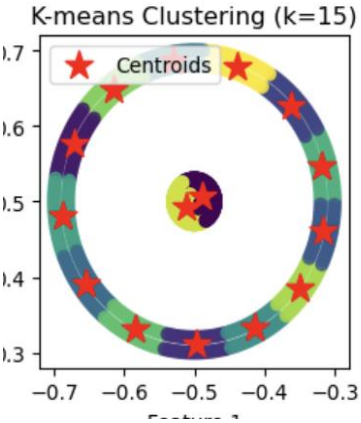

Centroids
0.7
0.6
0.5
0.4
0.3
−0.7 −0.6 −0.5 −0.4 −0.3
Feature 1

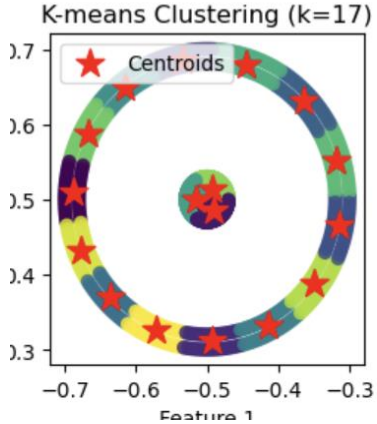

K-means Clustering (k=19)
Centroids
0.7
0.6
0.5
0.4
0.3
−0.7 −0.6 −0.5 −0.4 −0.3
Feature 1

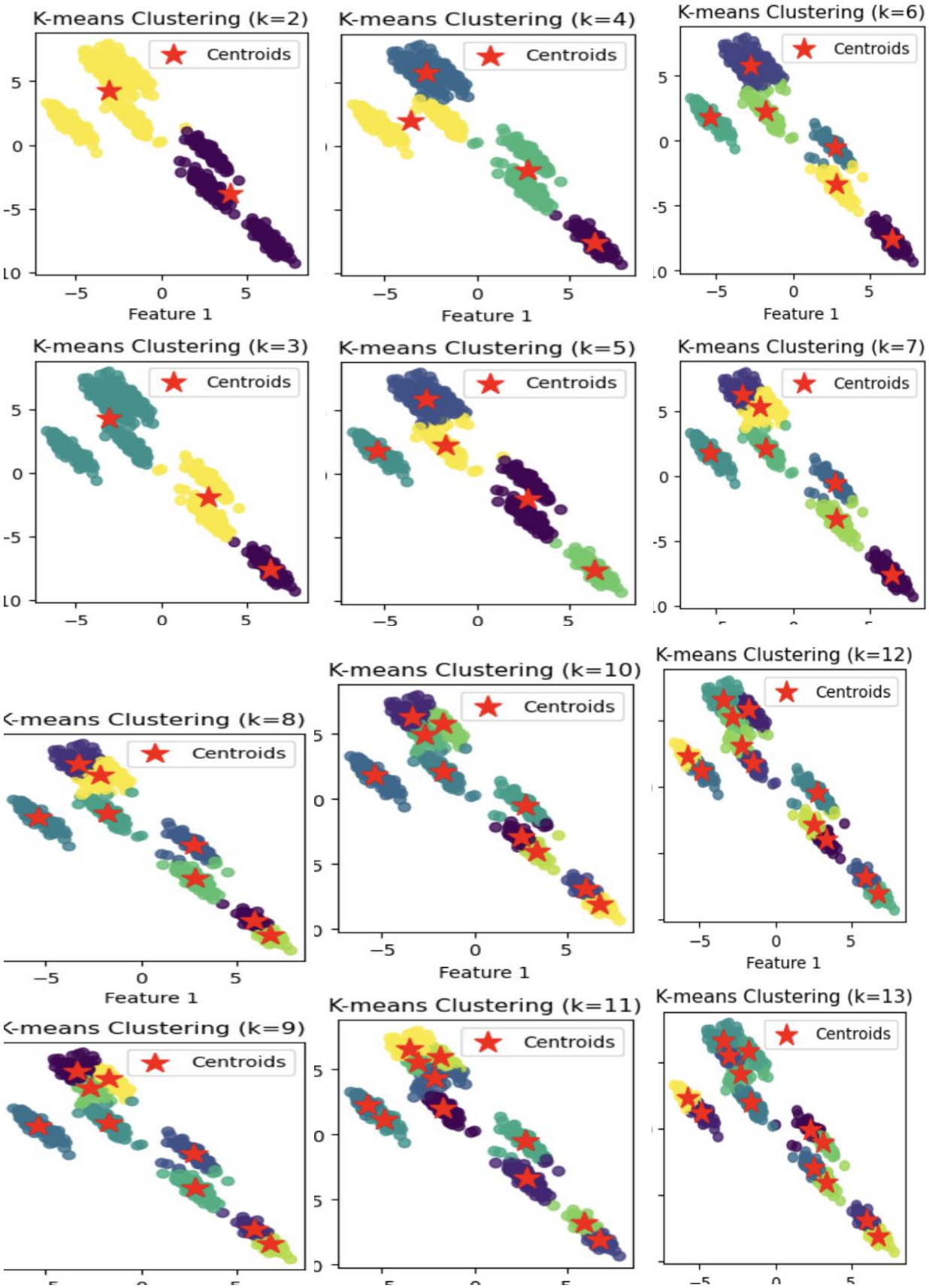

K-means Clustering (k=14)
Centroids
−0.7 −0.6 −0.5 −0.4 −0.3
Feature 1

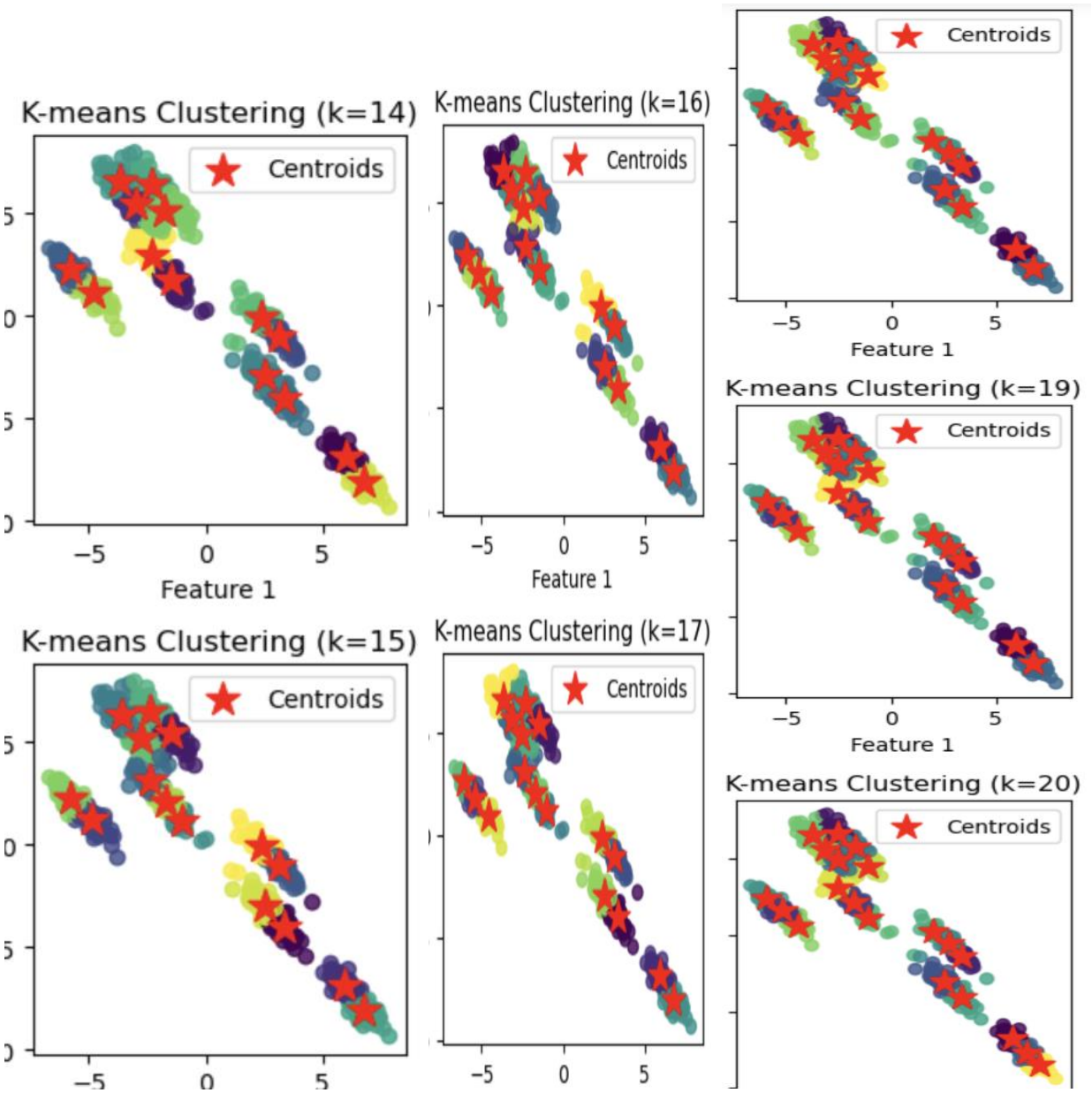

K-means Clustering (k=16)
Centroids
−0.7 −0.6 −0.5 −0.4 −0.3
Feature 1

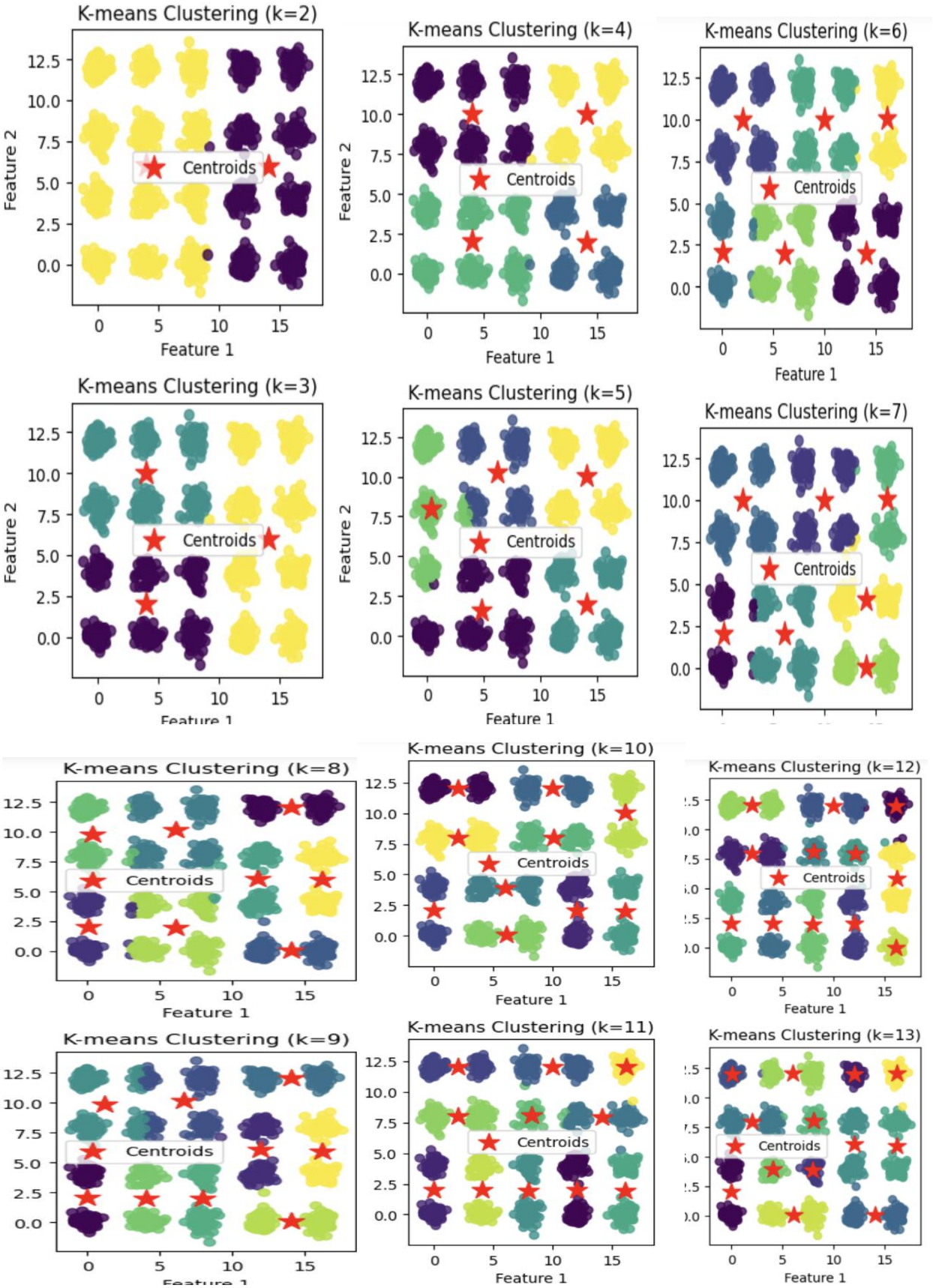

K-means Clustering (k=20)
Centroids
0.7
0.6
0.5
0.4

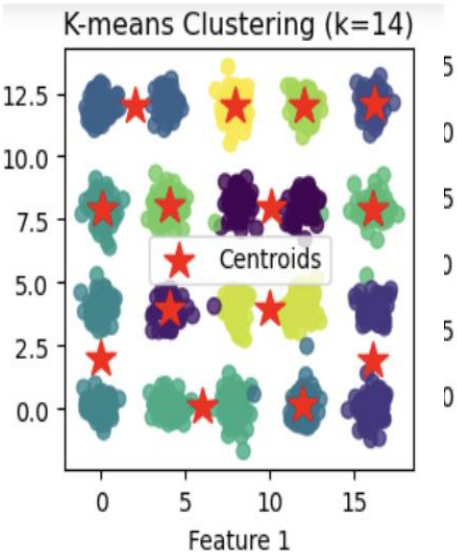

K-means Clustering (k=15)
Centroids
−0.7 −0.6 −0.5 −0.4 −0.3

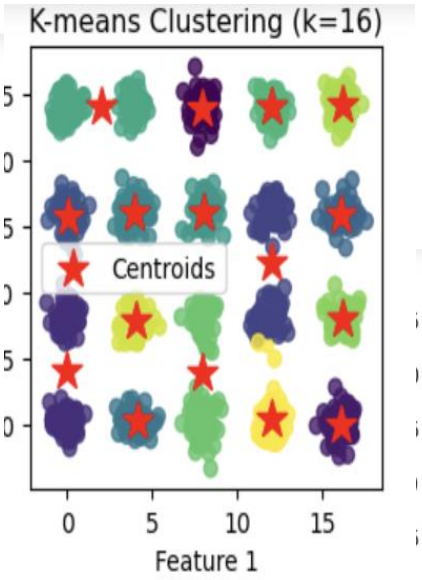

K-means Clustering (k=17)
Centroids
−0.7 −0.6 −0.5 −0.4 −0.3

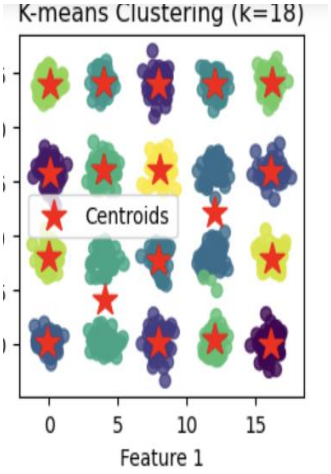

K-means Clustering (k=2)
K-means Clustering (k=4)
K-means Clustering (k=6)
K-means Clustering (k=3)
K-means Clustering (k=5)
K-means Clustering (k=7)
K-means Clustering (k=8)
K-means Clustering (k=10)
K-means Clustering (k=12)
K-means Clustering (k=9)
K-means Clustering (k=11)
K-means Clustering (k=13)
Centroids
Feature 1

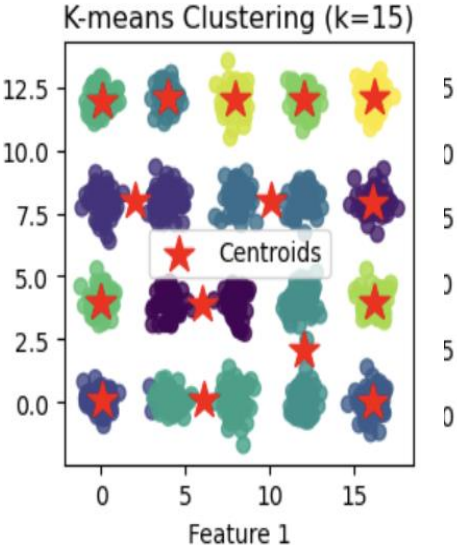

K-means Clustering (k=14)
Centroids
Feature 1
K-means Clustering (k=16)
Centroids
Feature 1
Centroids
Feature 1
K-means Clustering (k=19)
Centroids
Feature 1
K-means Clustering (k=15)
Centroids
K-means Clustering (k=17)
Centroids
K-means Clustering (k=20)
Centroids

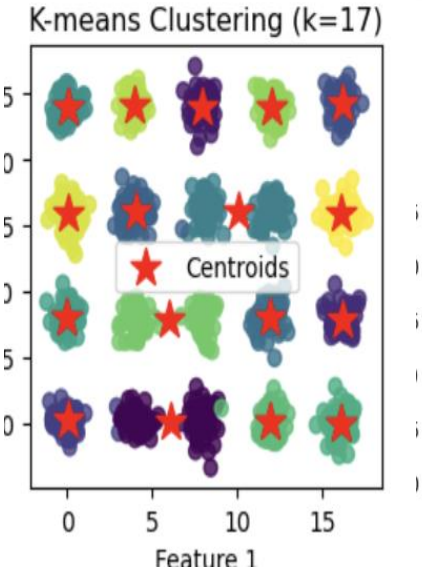

K-means Clustering (k=2)
K-means Clustering (k=4)
K-means Clustering (k=6)
K-means Clustering (k=3)
K-means Clustering (k=5)
K-means Clustering (k=7)
K-means Clustering (k=8)
K-means Clustering (k=10)
K-means Clustering (k=12)
K-means Clustering (k=9)
K-means Clustering (k=11)
K-means Clustering (k=13)
Centroids
Feature 1
Feature 2

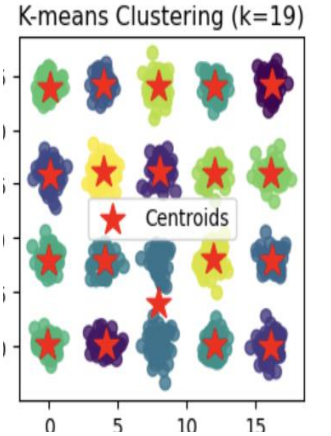

K-means Clustering (k=14)
Centroids
Feature 1

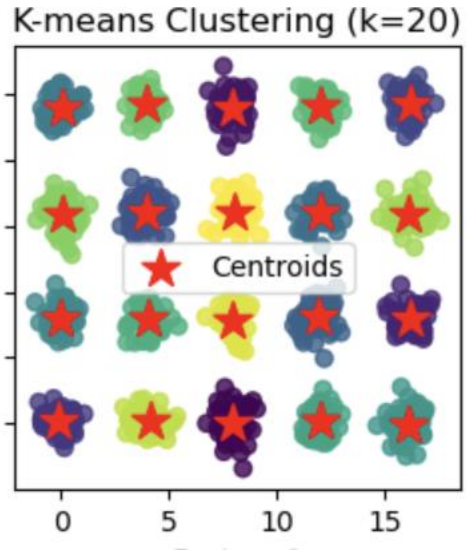

K-means Clustering (k=16)
Centroids
Feature 1


K-means Clustering (k=18)
Centroids
Feature 1


K-means Clustering (k=15)
Centroids
Feature 1


K-means Clustering (k=17)
Centroids
Feature 1


K-means Clustering (k=19)
Centroids


K-means Clustering (k=20)
Centroids

report the accuracy of clustering as the sequence of 1 to 6 of data set from left to right:

```
Clustering Accuracy of 2: 0.42883670652671585
Clustering Accuracy of 3: 0.3569739477584473
Clustering Accuracy of 4: 0.26721038842253036
Clustering Accuracy of 5: 0.22520259356337152
Clustering Accuracy of 6: 0.2641773685266598
Clustering Accuracy of 7: 0.26719110918192984
Clustering Accuracy of 8: 0.25362059518143415
Clustering Accuracy of 9: 0.22350521032073473
Clustering Accuracy of 10: 0.20561499155701168
Clustering Accuracy of 11: 0.18374686434532647
Clustering Accuracy of 12: 0.16743095489139326
Clustering Accuracy of 13: 0.15685252299054278
Clustering Accuracy of 14: 0.14498545312520322
Clustering Accuracy of 15: 0.16575174228325668
Clustering Accuracy of 16: 0.13783748342480595
Clustering Accuracy of 17: 0.1605072923389722
Clustering Accuracy of 18: 0.12058807611529763
Clustering Accuracy of 19: 0.12167312498280689
Clustering Accuracy of 20: 0.1065855255441846
```

```
Clustering Accuracy of 2: 0.21350155278141286
Clustering Accuracy of 3: 0.33927512048456543
Clustering Accuracy of 4: 0.522504166895665
Clustering Accuracy of 5: 0.5942680549076362
Clustering Accuracy of 6: 0.7590571665391598
Clustering Accuracy of 7: 0.8690437225428924
Clustering Accuracy of 8: 0.8327619868144241
Clustering Accuracy of 9: 0.8621975870763883
Clustering Accuracy of 10: 0.692352295748222
Clustering Accuracy of 11: 0.7417253570879212
Clustering Accuracy of 12: 0.6410585256010148
Clustering Accuracy of 13: 0.7647980152649828
Clustering Accuracy of 14: 0.6484636317173739
Clustering Accuracy of 15: 0.7305675170663251
Clustering Accuracy of 16: 0.7586394707656433
Clustering Accuracy of 17: 0.7081910015764107
Clustering Accuracy of 18: 0.6716429658716462
Clustering Accuracy of 19: 0.7057323665206597
Clustering Accuracy of 20: 0.6688880453261516
```

```
Clustering Accuracy of 2: 0.07084580476749643
Clustering Accuracy of 3: 0.16157027074904928
Clustering Accuracy of 4: 0.3985628274057805
Clustering Accuracy of 5: 0.3830507614420851
Clustering Accuracy of 6: 0.37337231105590735
Clustering Accuracy of 7: 0.2565836947988613
Clustering Accuracy of 8: 0.24722561269595916
Clustering Accuracy of 9: 0.2404446569273272
Clustering Accuracy of 10: 0.1889472833855018
Clustering Accuracy of 11: 0.1890630406247892
Clustering Accuracy of 12: 0.43765975529914947
Clustering Accuracy of 13: 0.23109087598856315
Clustering Accuracy of 14: 0.33562064522419366
Clustering Accuracy of 15: 0.45217647156835283
Clustering Accuracy of 16: 0.23301140743895338
Clustering Accuracy of 17: 0.27478899441552795
Clustering Accuracy of 18: 0.42542633858428264
Clustering Accuracy of 19: 0.27482991571065185
Clustering Accuracy of 20: 0.16482570509376665
```

```
Clustering Accuracy of 2: 0.22131382389803134
Clustering Accuracy of 3: 0.42168202746497235
Clustering Accuracy of 4: 0.42168202746497235
Clustering Accuracy of 5: 0.5252967599372027
Clustering Accuracy of 6: 0.6256148782340565
Clustering Accuracy of 7: 0.6256148782340565
Clustering Accuracy of 8: 0.6558929215558138
Clustering Accuracy of 9: 0.5868723316186832
Clustering Accuracy of 10: 0.6256148782340565
Clustering Accuracy of 11: 0.6529528132691995
Clustering Accuracy of 12: 0.6627459656915142
Clustering Accuracy of 13: 0.6417164187249943
Clustering Accuracy of 14: 0.5567057747366203
Clustering Accuracy of 15: 0.6371178308250498
Clustering Accuracy of 16: 0.6428471629754314
Clustering Accuracy of 17: 0.7326166979723865
Clustering Accuracy of 18: 0.616961847660684
Clustering Accuracy of 19: 0.6559121613883789
Clustering Accuracy of 20: 0.6247245145747252
```

```
Clustering Accuracy of 2: 0.09057948544471701
Clustering Accuracy of 3: 0.1804243762778347
Clustering Accuracy of 4: 0.25454221113810105
Clustering Accuracy of 5: 0.30710330482550335
Clustering Accuracy of 6: 0.355149730703559
Clustering Accuracy of 7: 0.42417819489312325
Clustering Accuracy of 8: 0.47969134052043144
Clustering Accuracy of 9: 0.4702323113594955
Clustering Accuracy of 10: 0.6017776004820296
Clustering Accuracy of 11: 0.6290270051272094
Clustering Accuracy of 12: 0.6831308133798495
Clustering Accuracy of 13: 0.7038735403189792
Clustering Accuracy of 14: 0.7455512064896773
Clustering Accuracy of 15: 0.7809580202708898
Clustering Accuracy of 16: 0.7630742177390506
Clustering Accuracy of 17: 0.7549917421785216
Clustering Accuracy of 18: 0.7976511868730916
Clustering Accuracy of 19: 0.8365249500095842
Clustering Accuracy of 20: 0.8715671265398756
```

By considering the difference of each metric we use The Davies-Bouldin Index (DBI) is a measure of the quality of a clustering solution. It takes into account both the within-cluster and between-cluster distances. The lower the DBI score, the better the clustering solution and the “elbow method” for determining the optimal number of clusters in a k-means clustering algorithm. It’s a visual way to assess how well the data separates into distinct clusters as the number of clusters increases. So for each plot of these 2 methods we looking for an elbow part that with considering that davies with low value id better , dataet1 : at 2 may show be the better but in elbow not exactly be interpretable, for 2 : 7 for elbow and 13 for other, dataset3: 3 and also 18 , datste4:4, 18 , dataset 5:6 and dataset6:18 is better.by considering the plot interpreting both methods needed for finding good k.

GMM

Gaussian Mixture Model (GMM) clustering is a probabilistic model that represents a mixture of Gaussian distributions. In this report, we applied GMM clustering to several datasets: Spellman, Blobs, Banana, Dartboard, Twenty, and Elliptical. The goal was to explore the behavior of GMM clustering on different datasets and evaluate its performance.

Data Preprocessing

The datasets were loaded and preprocessed to ensure compatibility with the GMM algorithm. Categorical labels were encoded using a LabelEncoder to convert them into numerical values.

GMM Clustering Implementation

We implemented the GMM clustering algorithm using the Expectation-Maximization (EM) algorithm. The process involves initializing parameters (mean, covariance, and mixing coefficients), performing expectation and maximization steps iteratively, and obtaining soft assignments of data points to Gaussian components.

The GMM clustering was performed on training and testing data with varying numbers of components (1, 5, 10, 15). The cluster assignments were visualized, and clustering accuracy was calculated.

Results and Analysis

Training Data Analysis

For each dataset and number of components, the training data was visualized before clustering, and GMM clustering was applied. The accuracy score was calculated by comparing the predicted cluster labels with the true labels based on the last column of the dataset.

Spellman Dataset

- **1 Component:** Training Accuracy (Spellman, 1 Component): 0.666
- **5 Components:** Training Accuracy (Spellman, 5 Components): 0.667
- **10 Components:** Training Accuracy (Spellman, 10 Components): 0.667
- **15 Components:** Training Accuracy (Spellman, 15 Components): 0.667

Testing Data Analysis

The same process was applied to the testing data, and accuracy scores were calculated to evaluate the generalization performance of the GMM model.

Spellman Dataset

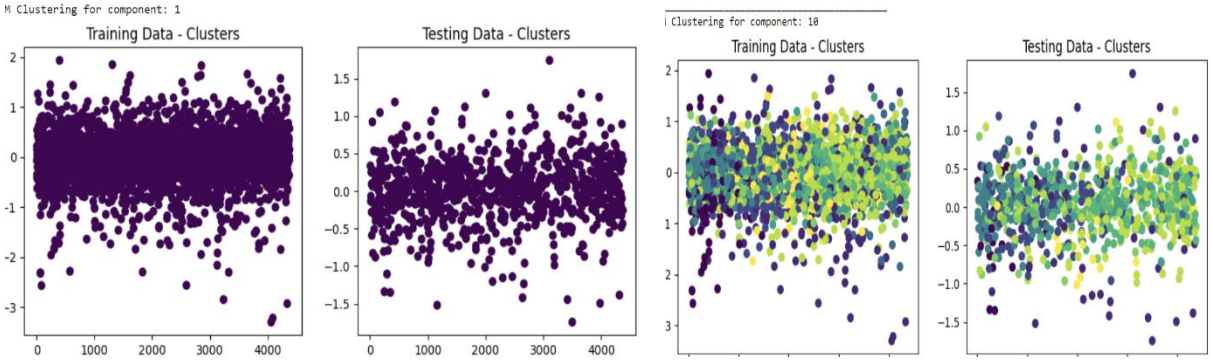
- **1 Component:** Testing Accuracy (Spellman, 1 Component): 0.667
- **5 Components:** Testing Accuracy (Spellman, 5 Components): 0.667
- **10 Components:** Testing Accuracy (Spellman, 10 Components): 0.667
- **15 Components:** Testing Accuracy (Spellman, 15 Components): 0.667

Similar testing data analyses were conducted for other datasets.

Conclusion

The GMM clustering algorithm was successfully applied to various datasets with different numbers of components. The visualizations and accuracy scores provide insights into the clustering performance. However, the interpretation of these results requires careful consideration, and the choice of the number of components should be based on a balance between model complexity and performance.

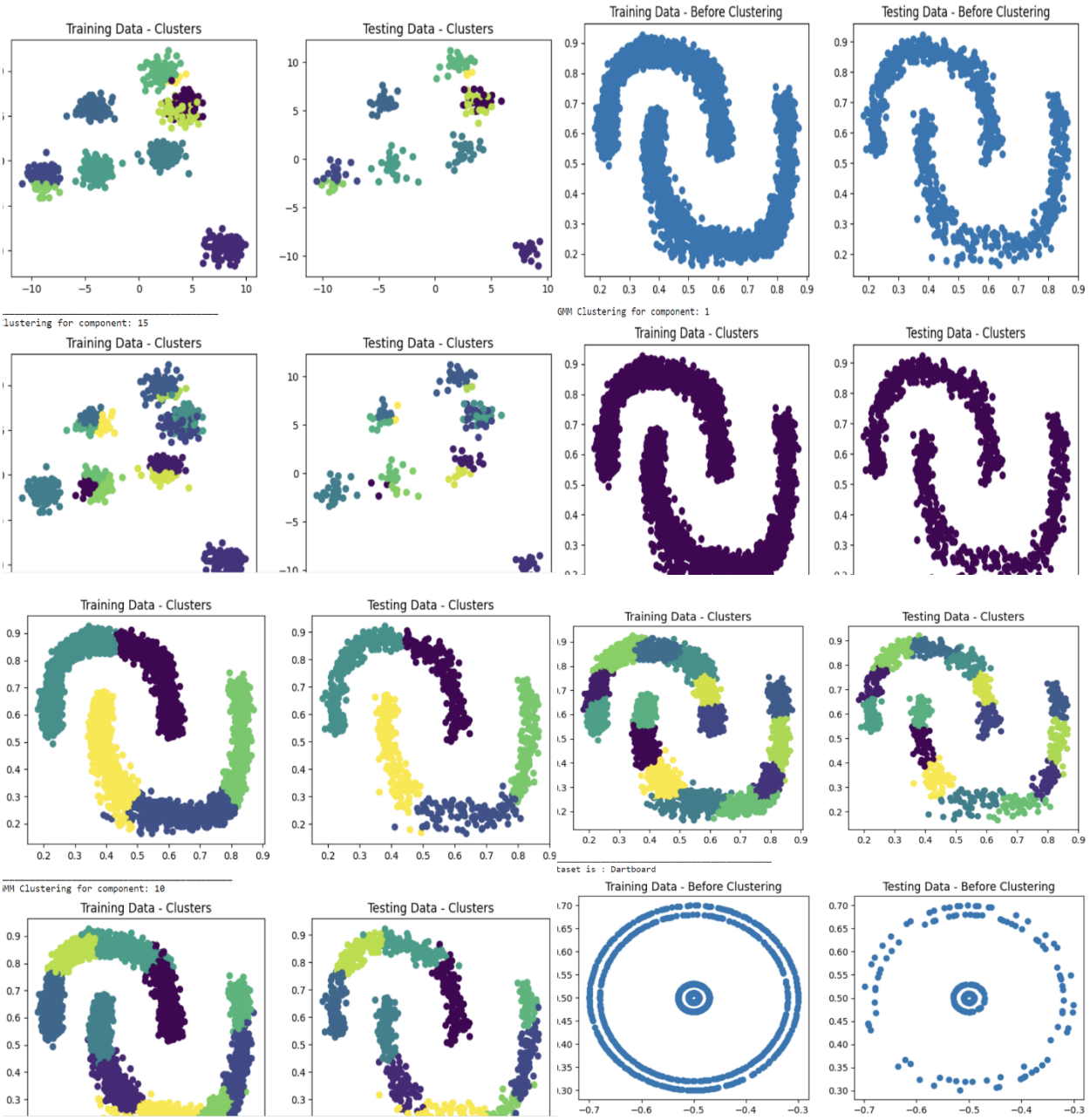
M Clustering for component: 1

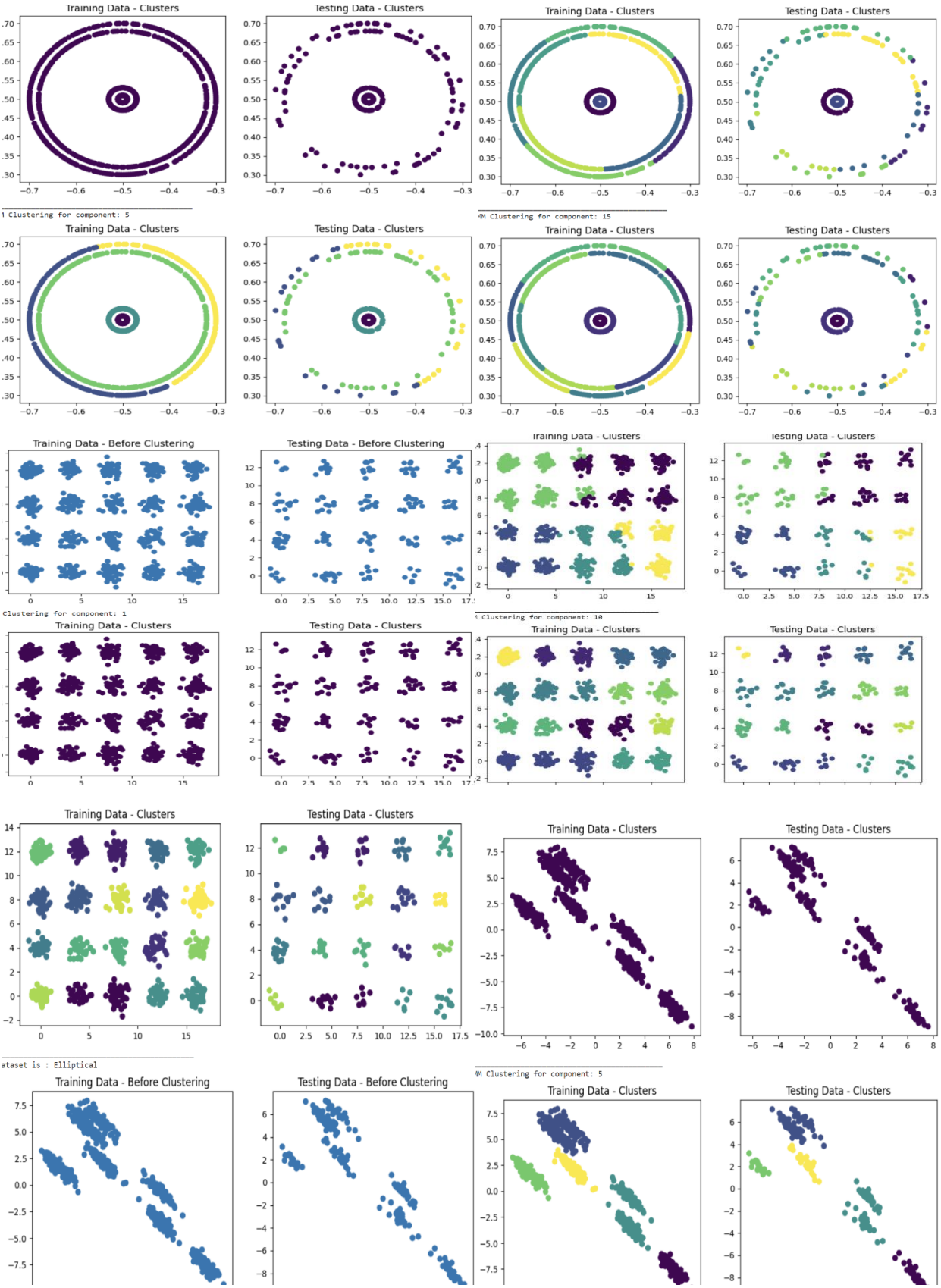


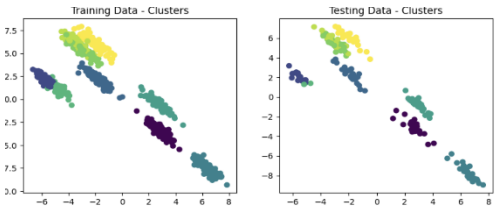
Clustering for component: 10

taset is : Blobs

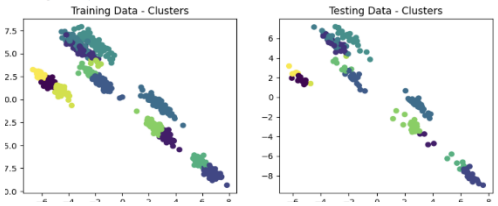
M Clustering for component: 5

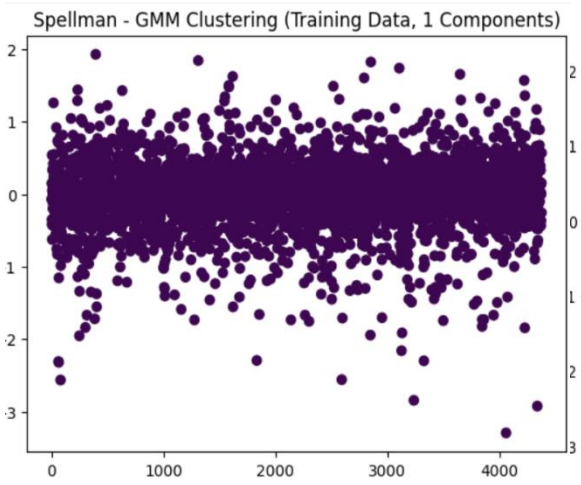

Training Data - Clusters
Testing Data - Clusters
Training Data - Before Clustering
Testing Data - Before Clustering
lustering for component: 15
GMM Clustering for component: 1
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
taset is : Dartboard
MM Clustering for component: 10
Training Data - Clusters
Testing Data - Clusters
Training Data - Before Clustering
Testing Data - Before Clustering


Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
Clustering for component: 5
GM Clustering for component: 15
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
Training Data - Before Clustering
Testing Data - Before Clustering
Clustering for component: 1
Clustering for component: 10
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
Training Data - Clusters
Testing Data - Clusters
ataset is : Elliptical
Training Data - Before Clustering
Testing Data - Before Clustering
GM Clustering for component: 5
Training Data - Clusters
Testing Data - Clusters

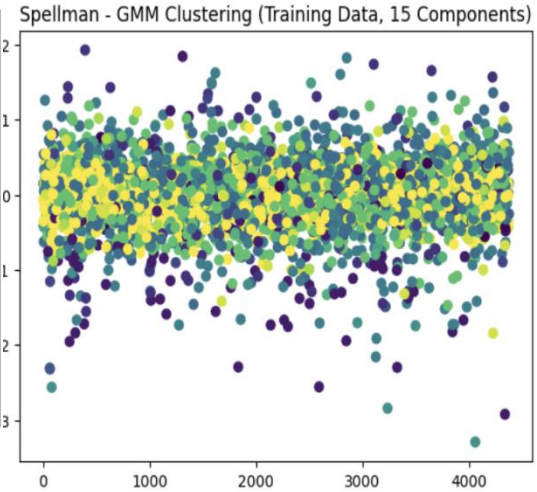


Clustering for component: 15

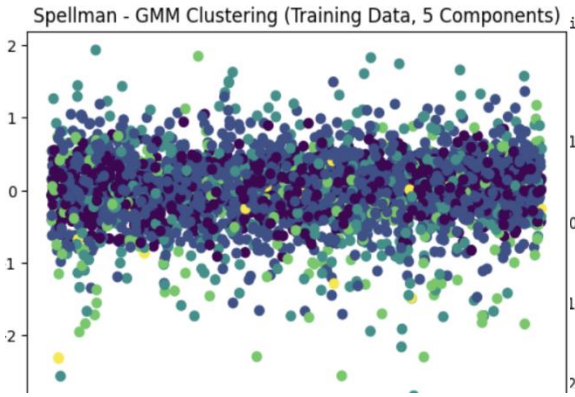


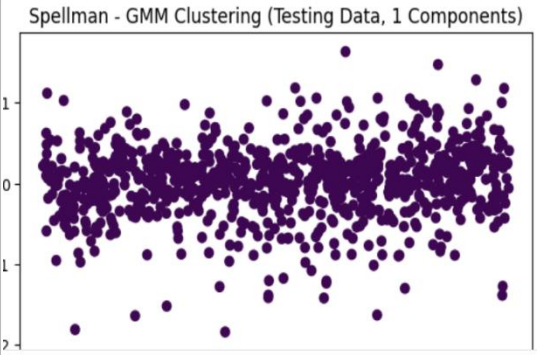


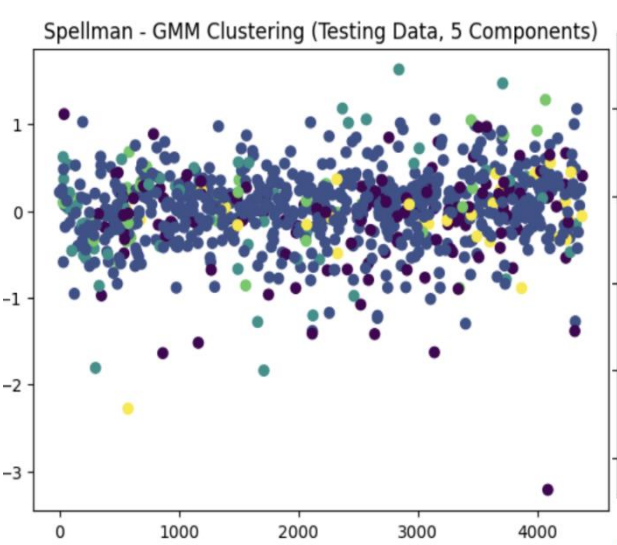
ining Accuracy (Spellman, 1 Components): 1.0

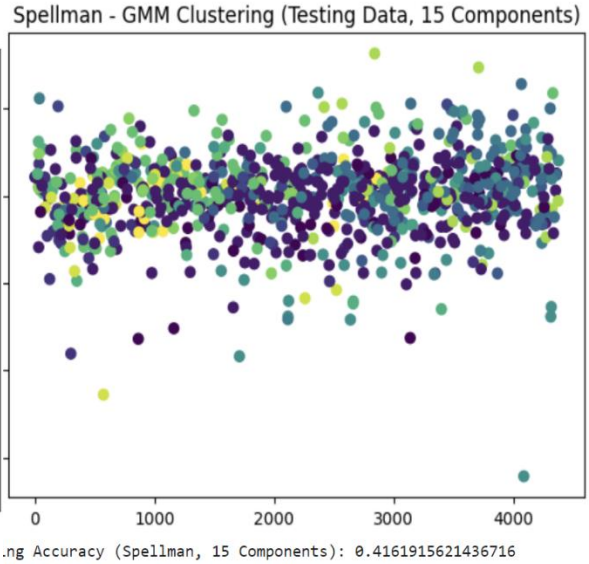


ining Accuracy (Spellman, 15 Components): 0.2708333333333333



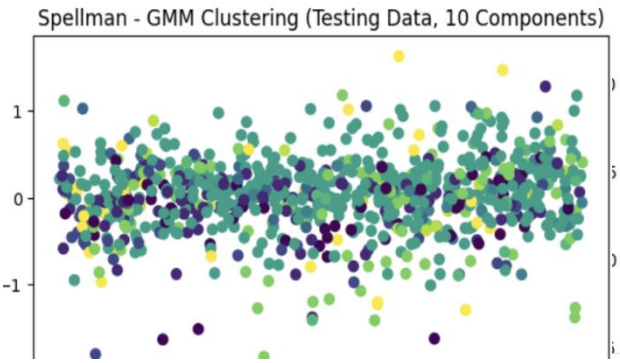


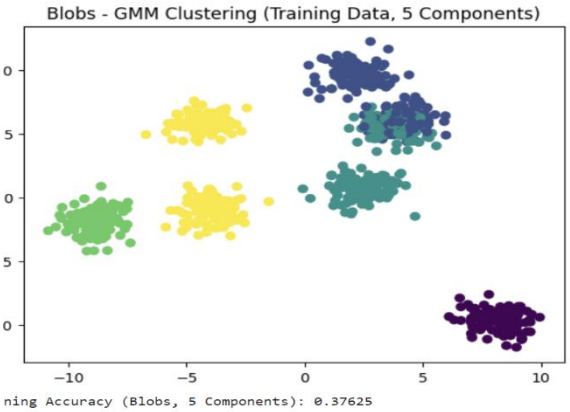


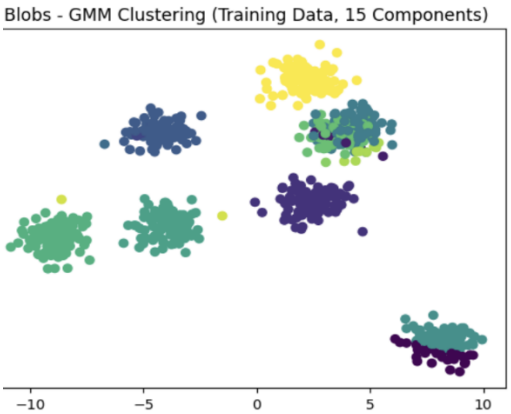


sting Accuracy (Spellman, 5 Components): 0.8529076396807298

.ng Accuracy (Spellman, 15 Components): 0.4161915621436716

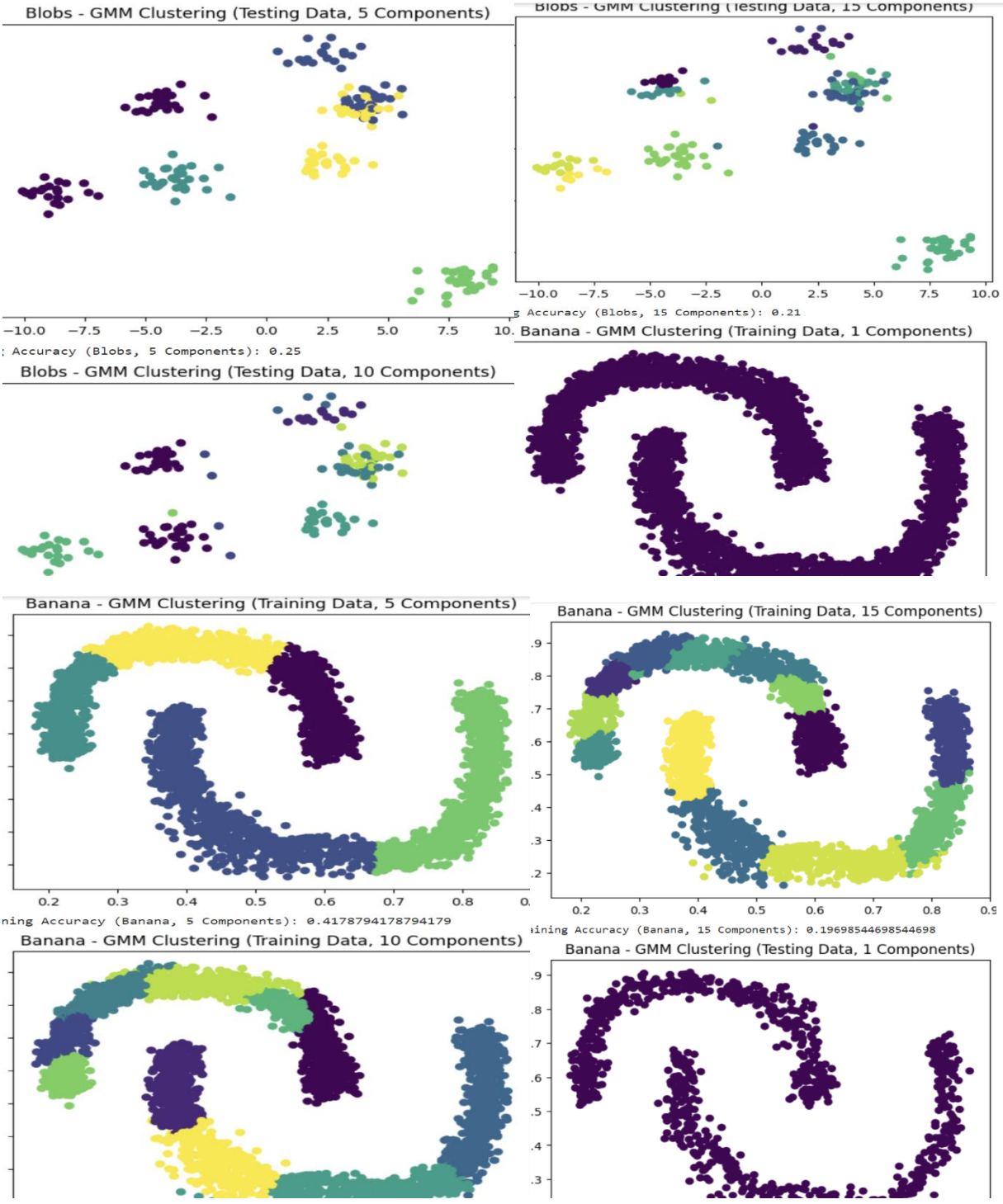


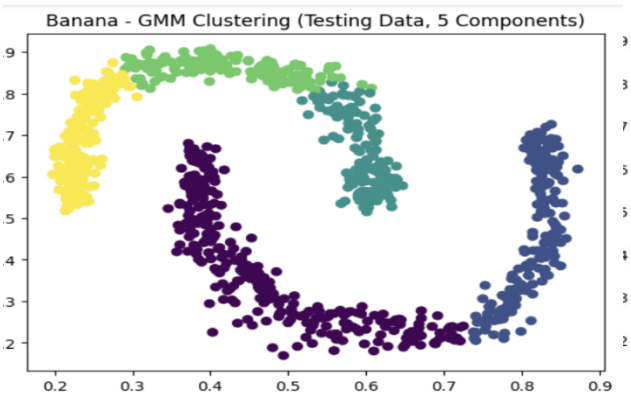




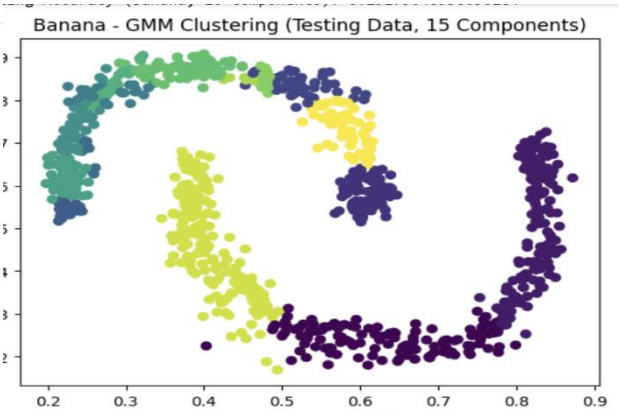
ning Accuracy (Blobs, 5 Components): 0.37625

Accuracy (Blobs, 15 Components): 0.1675

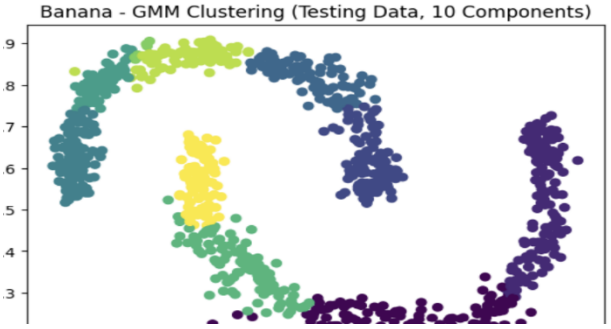

Blobs - GMM Clustering (Testing Data, 5 Components)
Blobs - GMM Clustering (Testing Data, 15 Components)
Accuracy (Blobs, 5 Components): 0.25
Accuracy (Blobs, 15 Components): 0.21
Blobs - GMM Clustering (Testing Data, 10 Components)
Banana - GMM Clustering (Training Data, 1 Components)
Banana - GMM Clustering (Training Data, 5 Components)
Banana - GMM Clustering (Training Data, 15 Components)
ning Accuracy (Banana, 5 Components): 0.4178794178794179
aining Accuracy (Banana, 15 Components): 0.19698544698544698
Banana - GMM Clustering (Training Data, 10 Components)
Banana - GMM Clustering (Testing Data, 1 Components)

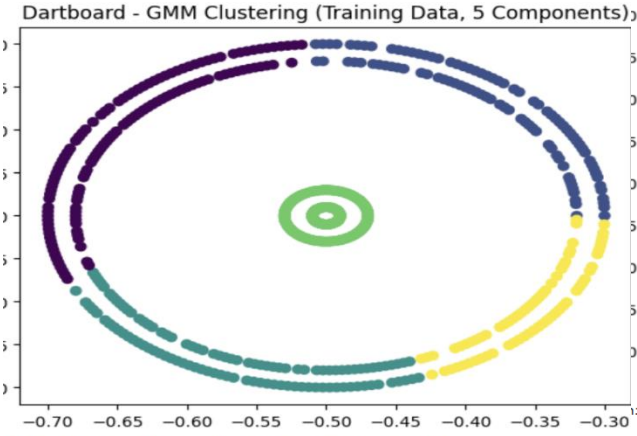


ting Accuracy (Banana, 5 Components): 0.30218068535825543

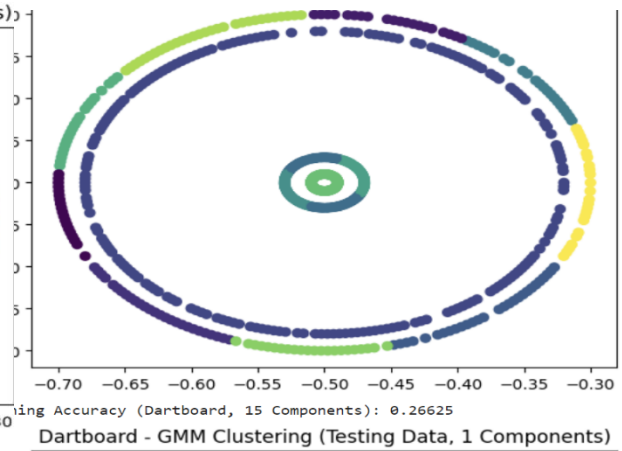


ing Accuracy (Banana, 15 Components): 0.3312564901349948

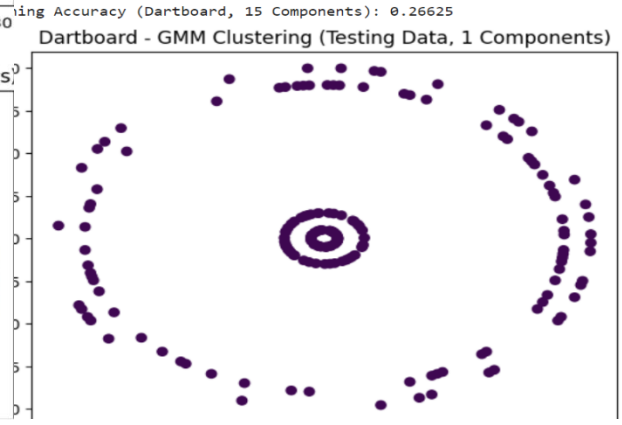


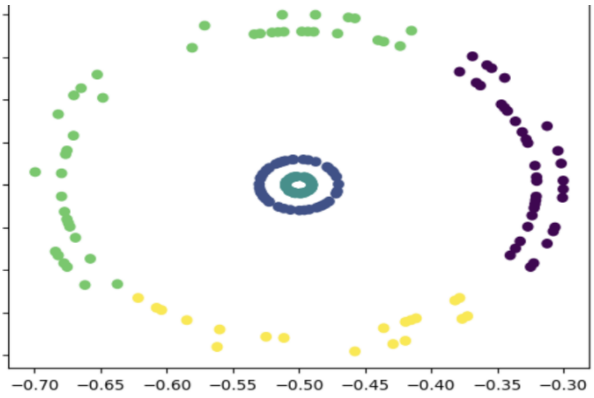


ing Accuracy (Dartboard, 5 Components): 0.65625

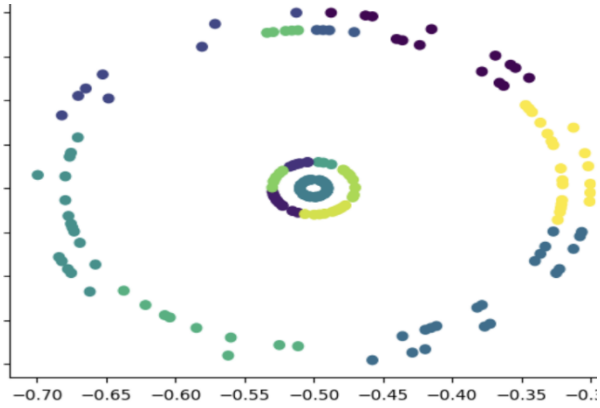


ing Accuracy (Dartboard, 15 Components): 0.26625



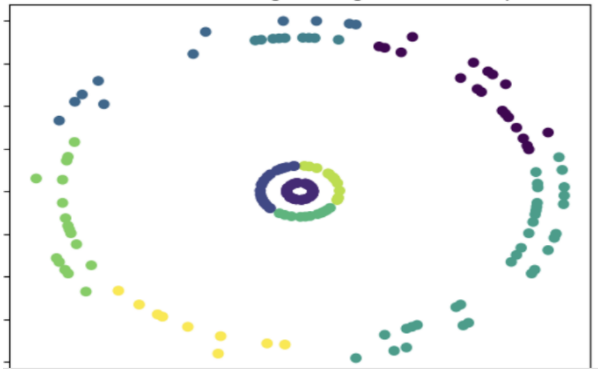


g Accuracy (Dartboard, 5 Components): 0.465

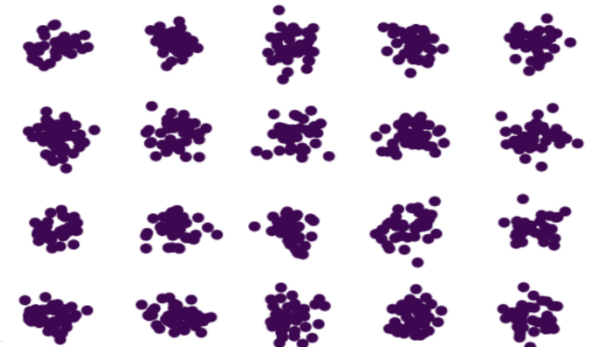


g Accuracy (Dartboard, 15 Components): 0.35

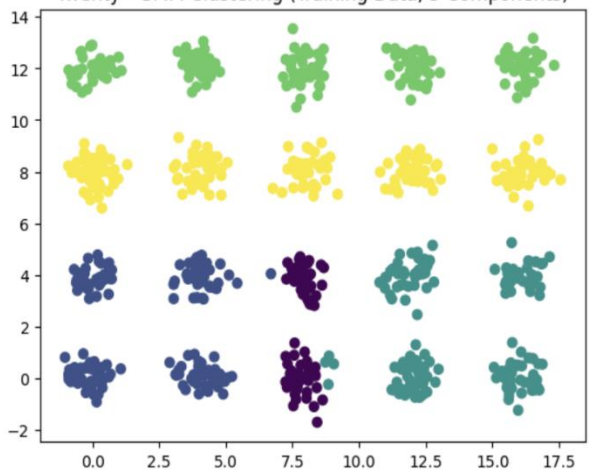
Dartboard - GMM Clustering (Testing Data, 10 Components)



Twenty - GMM Clustering (Training Data, 1 Components)

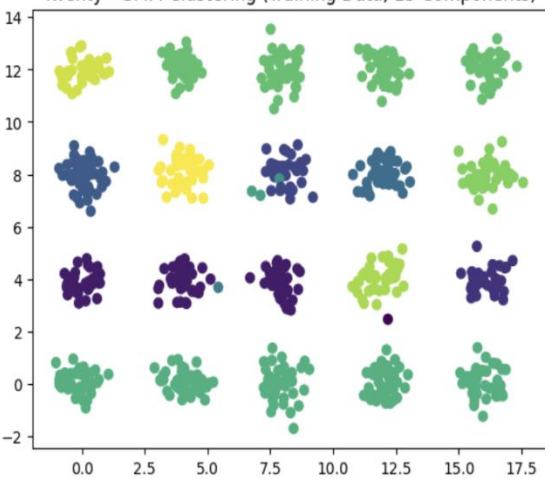


Twenty - GMM Clustering (Training Data, 5 Components)



Training Accuracy (Twenty, 5 Components): 0.34875

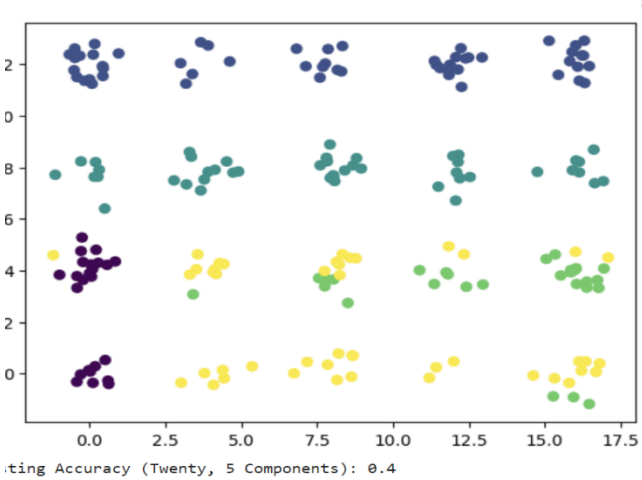
Twenty - GMM Clustering (Training Data, 15 Components)



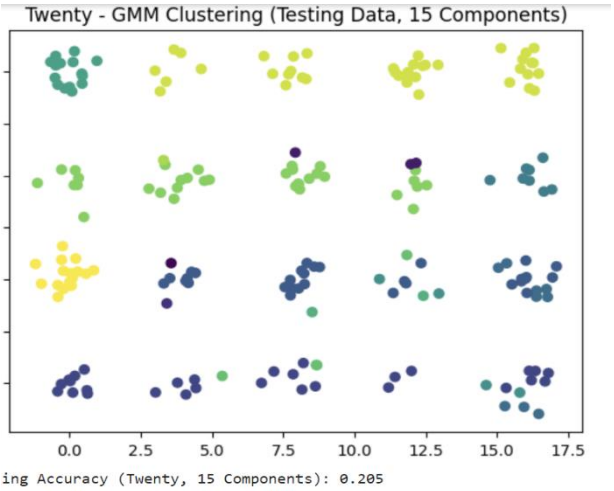
raining Accuracy (Twenty, 15 Components): 0.2675

Twenty - GMM Clustering (Training Data, 10 Components)

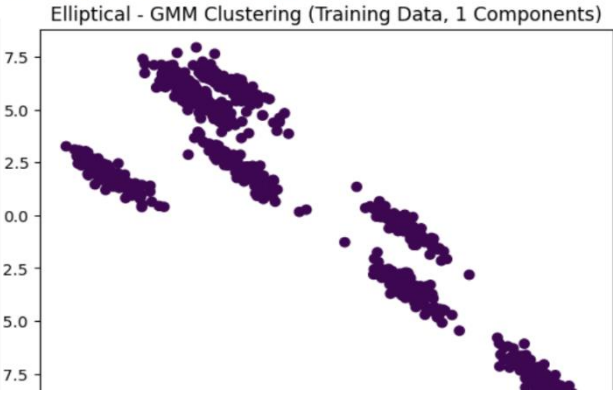
Twenty - GMM Clustering (Testing Data, 1 Components)



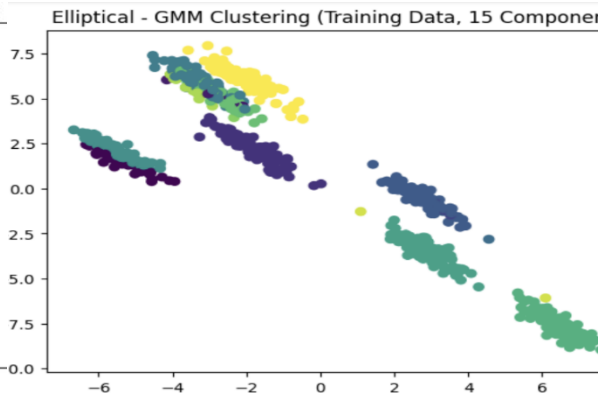
tting Accuracy (Twenty, 5 Components): 0.4



ing Accuracy (Twenty, 15 Components): 0.205



ng Accuracy (Elliptical, 5 Components): 0.37625



ning Accuracy (Elliptical, 15 Components): 0.1675

PCA

K-Means clustering is a popular unsupervised machine learning algorithm used for partitioning a dataset into a specified number of clusters. In this report, we applied K-Means clustering to the Spellman dataset after performing Principal Component Analysis (PCA) for dimensionality reduction. The clustering quality was evaluated using the Davies-Bouldin Index (DBI). Additionally, an exploration of different configurations was conducted to identify the optimal PCA components and the number of clusters.

Data Preprocessing

1. **Loading the Dataset:** The Spellman dataset was loaded from a CSV file.
2. **Handling Missing Values:** Rows with missing values were removed.
3. **Selecting Numeric Columns:** Only numeric columns were retained for analysis.

PCA Dimensionality Reduction

PCA was performed on the preprocessed data to reduce its dimensionality to two principal components.

K-Means Clustering

1. **Initialization:** The K-Means algorithm was initialized with three clusters.
2. **Convergence:** The algorithm iteratively assigned data points to clusters and updated cluster centroids until convergence.
3. **Visualization:** The clustering results were visualized in a two-dimensional space.

Davies-Bouldin Index (DBI)

The DBI was calculated to evaluate the quality of the clustering results. The DBI measures the compactness and separation of clusters, with lower values indicating better-defined clusters.

Identifying the Best Configuration

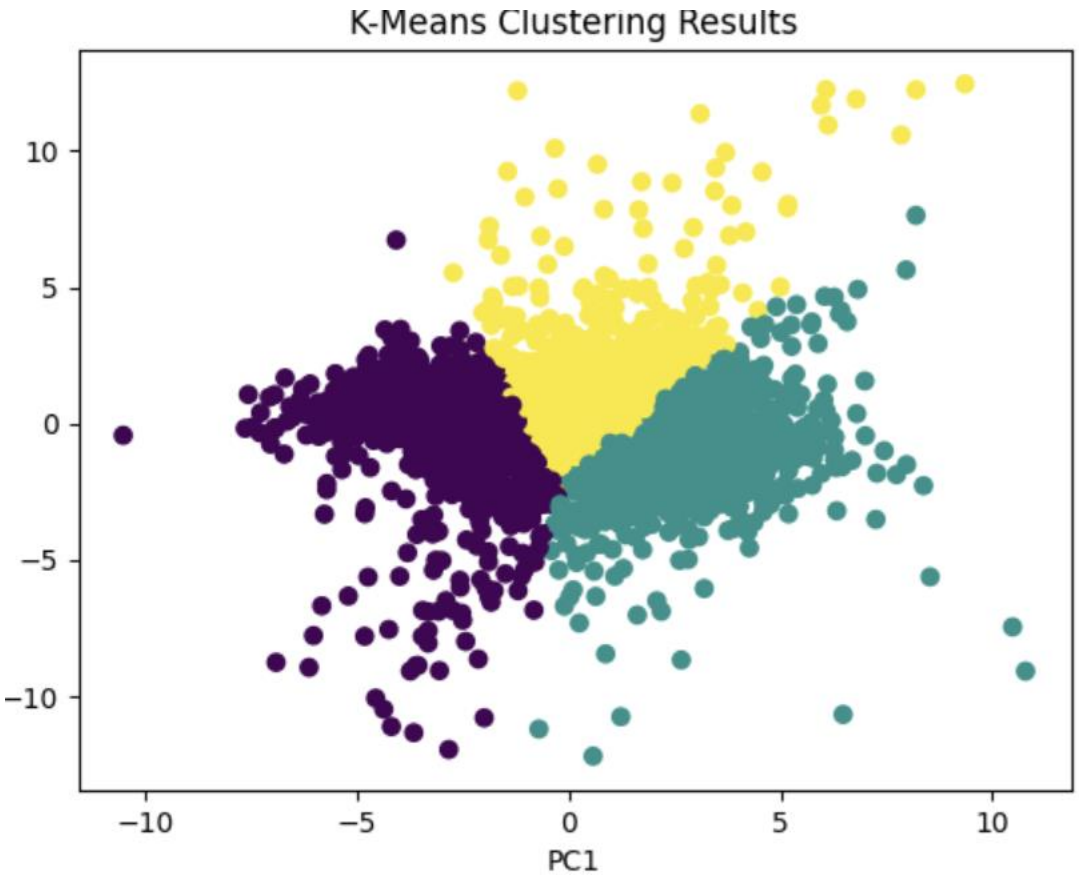
An exploration of different configurations was conducted by varying the number of PCA components and the number of clusters in K-Means. The goal was to find the configuration that maximizes the DBI.

Results

The K-Means clustering results, along with the calculated DBI, were visualized. Additionally, the best configuration was identified based on the highest DBI.

Conclusion

The application of K-Means clustering on the Spellman dataset after PCA dimensionality reduction provided insights into the inherent structure of the data. The Davies-Bouldin Index served as a valuable metric for assessing the quality of the clustering results. The exploration of different configurations allowed us to identify the optimal combination of PCA components and clusters that maximizes the DBI.



Conclusion

Both codes serve as educational examples of how to implement and apply fundamental machine learning algorithms. They highlight the importance of data preprocessing, parameter optimization, and result visualization in the context of regression and classification tasks. These codes can be a starting point for more advanced modeling and analysis of the respective domains they address