

Clustering Results Report

Introduction

In this project, we applied multiple clustering techniques to segment the customers into distinct groups. The following techniques were employed:

- **Agglomerative Clustering**
- **KMeans Clustering**
- **Affinity Propagation**
- **Self-Organizing Maps (SOM)**

These techniques were chosen to explore various clustering strategies and determine which approach best segments the customer data.

Clustering Techniques Used

1. **Agglomerative Clustering:**

Agglomerative clustering is a bottom-up hierarchical clustering method. Initially, each data point is considered its own cluster, and clusters are merged iteratively based on a distance metric (e.g., Euclidean distance) until all points belong to a single cluster.

2. **KMeans Clustering:**

KMeans is a partitional clustering technique that divides the dataset into K predefined clusters. It works by minimizing the variance within each cluster and using the centroid as the cluster center. The number of clusters must be specified before running the algorithm.

3. **Affinity Propagation:**

Affinity Propagation is a clustering algorithm that works by finding “exemplars” (representative data points) and assigning all other points to the nearest exemplar. Unlike KMeans, it does not require the number of clusters to be specified in advance.

4. **Self-Organizing Maps (SOM):**

SOM is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically 2D) representation of the input data. It is used for clustering by mapping high-dimensional data into a grid of clusters.

Clustering Results

- **Agglomerative Clustering:**
 - Number of Clusters: 5

- **DB Index:** 0.9088
 - The DB index for Agglomerative Clustering indicates a relatively good clustering structure, suggesting the clusters formed are distinct.
 - **KMeans Clustering:**
 - Number of Clusters: 5
 - **DB Index:** 0.9141
 - KMeans produced comparable results with a slightly higher DB index, indicating slightly better-defined clusters compared to Agglomerative Clustering.
 - **Affinity Propagation:**
 - **DB Index:** 1.0199
 - Affinity Propagation had a higher DB index, which might indicate less distinct or well-separated clusters compared to the other methods.
 - **Self-Organizing Maps (SOM):**
 - Number of Clusters: 2
 - **DB Index:** 0.7494
 - SOM produced clusters that seem less distinct, as indicated by the relatively lower DB index. It may be due to the complexity of the SOM algorithm, which tends to produce less well-defined clusters in high-dimensional datasets.
-

Cluster Evaluation

The **Davies-Bouldin (DB) Index** is a metric used to evaluate the quality of clusters, with lower values indicating better-defined and more distinct clusters. Based on the DB index, the best clustering technique for this dataset appears to be **KMeans** with a DB index of **0.9141**.

The **Affinity Propagation** method, while not as effective in producing well-separated clusters, might still provide useful insights in certain cases where the number of clusters is unknown.

Conclusion

- **Best Technique:** Based on the DB index, **KMeans** clustering offers the most well-defined clusters among the techniques used.

- **Technique Choice:** The choice of clustering algorithm is highly dependent on the nature of the dataset. KMeans is most useful when the number of clusters is known, while Affinity Propagation can be useful when the number of clusters is not predefined.