Clustering Results Report

1. Overview

This report summarizes the results of the clustering analysis performed using the KMeans algorithm. The goal was to determine the optimal number of clusters (k) and evaluate the clustering quality using the metric Davies-Bouldin Index (DBI). Additionally, Principal Component Analysis (PCA) with 2 components was used for visualization purposes.

2. Clustering Metrics

The following table summarizes the clustering metrics for different values of k:

k	DB Index	Silhouette Score
3	1.329780	0.250499
4	1.232277	0.270205
5	1.139725	0.278379
6	1.063615	0.292920
7	1.034679	0.285205
8	0.979509	0.282692
9	0.984836	0.303090

3. Optimal Number of Clusters

Based on the clustering metrics, the optimal number of clusters is determined as follows: The DBI measures the ratio of within-cluster scatter to between-cluster separation. Lower values indicate better clustering.

The minimum DBI is 0.979509 for k=8.

4. Other relevant clustering metrics.

Silhouette Score is used as the additional metric which has the highest value for k=9.

5. Visualization Using PCA:

To visualize the clusters, Principal Component Analysis (PCA) was applied to reduce the data to 2 components. The resulting 2D scatter plot provides an intuitive representation of the clusters. Below is an interpretation of the visualization:

6. Conclusion:

The clustering analysis successfully identified meaningful clusters in the data. The optimal number of clusters is 8, depending on the metric used. The visualization using PCA confirms that the clusters are reasonably well-separated and compact.