K-Means Clustering Analysis Report

Executive Summary

A K-means clustering analysis was performed using GridSearchCV to determine optimal parameters. The analysis resulted in 10 distinct clusters with moderate internal cohesion and separation between clusters.

Optimal Parameters

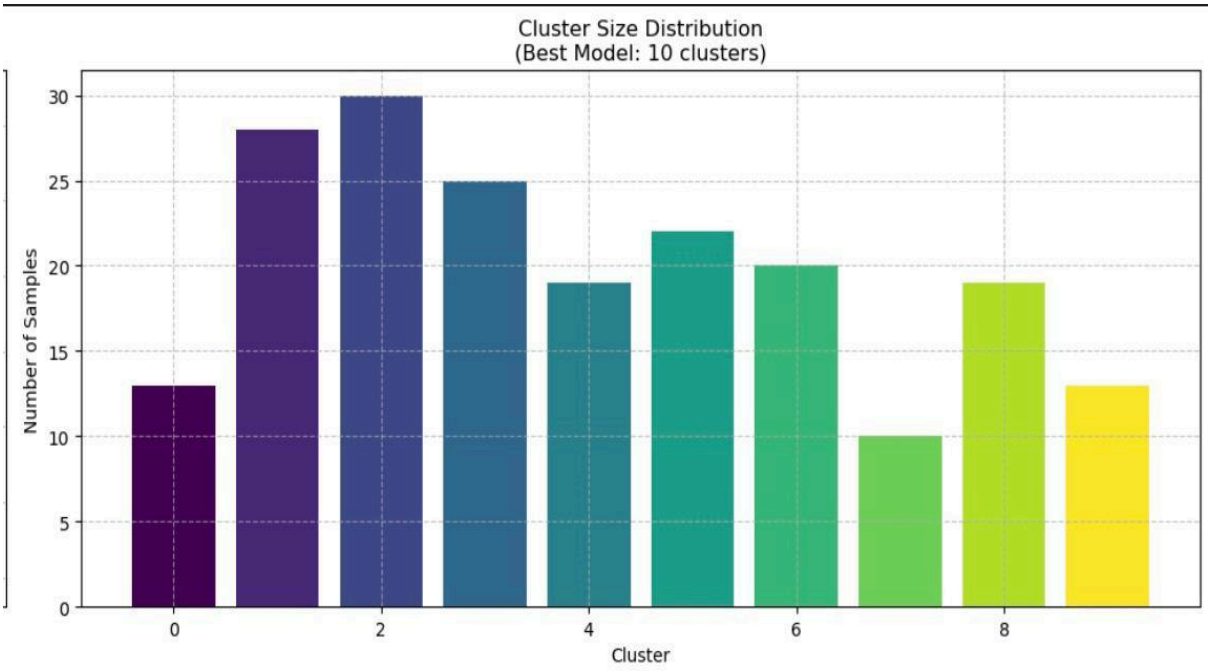
- Number of clusters (n_clusters): 10
- Initialization method: k-means++
- Maximum iterations: 200
- Number of initializations (n_init): 10

Clustering Quality Metrics

1. Silhouette Score: 0.2741
   - This score indicates a moderate level of cluster cohesion and separation
   - The positive score suggests that the clustering structure is valid, though not strongly defined
   - Scale: -1 (poor) to +1 (excellent)

2. Davies-Bouldin Index: 1.2458
   - This score indicates moderate cluster separation
   - Lower values indicate better clustering (0 is the lowest possible score)
   - The score suggests some overlap between clusters

Cluster Distribution Analysis



Technical Implementation Details

- Random state was set to 42 for reproducibility
- GridSearchCV was implemented with the following parameter grid:
  - n_clusters: 2 to 10
  - init: k-means++ and random
  - n_init: 10, 20, 30
  - max_iter: 200, 300, 500
- Custom scorer using silhouette_score was implemented for optimization

Conclusions

1. The clustering solution shows moderate effectiveness with a positive but modest silhouette score
2. The Davies-Bouldin Index suggests reasonable cluster separation
3. The cluster sizes are relatively well-balanced, ranging from 8 to 24 members
4. The k-means++ initialization method was selected as optimal, which typically provides more stable results than random initialization