Clustering Results Report

Harshavardhana Kandrakota

January 26, 2025

Abstract

This report presents the clustering analysis results conducted on the dataset. Key metrics such as the number of clusters formed, the Davies-Bouldin (DB) Index value, and other relevant clustering performance metrics are discussed in detail.

1 Introduction

Clustering is a vital unsupervised learning technique used for grouping similar data points. This analysis aims to identify meaningful patterns in the dataset and evaluate clustering performance using quantitative metrics.

2 Clustering Results

2.1 Number of Clusters Formed

The analysis resulted in [**X**] clusters. The number of clusters was determined using [methodology, e.g., the elbow method, silhouette analysis, etc.].

2.2 Davies-Bouldin Index

The Davies-Bouldin (DB) Index value for the clustering solution is [**Y**]. A lower DB Index indicates better-defined clusters, with values closer to zero being preferable.

2.3 Other Relevant Metrics

The following metrics were calculated to evaluate clustering performance:

- **Silhouette Score:** [Value]
  The silhouette score measures how similar an object is to its own cluster compared to other clusters. A score close to 1 indicates well-defined clusters.
- **Inertia:** [Value]
  Inertia quantifies the internal coherence of clusters. Lower values indicate compact and distinct clusters.
- **Calinski-Harabasz Index:** [Value]
  This index measures the ratio of the sum of between-cluster dispersion and within-cluster dispersion. Higher values indicate better clustering performance.

3 Discussion

The clustering results demonstrate [brief interpretation, e.g., "clear separation of data points into distinct groups" or "overlapping clusters due to high intra-cluster variance"]. Improvements can be achieved by [e.g., "tuning hyperparameters, using advanced clustering algorithms, or feature scaling"].

4 Conclusion

The clustering analysis identified **[X]** clusters, achieving a DB Index of **[Y]**. Additional metrics such as the silhouette score and Calinski-Harabasz Index further validate the quality of the clustering solution. Future work may include exploring alternative algorithms and feature engineering for enhanced clustering performance.

References

[1] Relevant references or methodologies used.