

Clustering Insights

To determine the optimal number of clusters, used WCSS(Within Cluster Sum of Squares), we can find the optimal number of clusters for the dataset. The plot of WCSS against the number of clusters will show the elbow point where the rate decrease slows down.

The model was run with K-Means clustering with 4 clusters. The Davies-Bouldin Index is used to evaluate the clustering quality. A value closer to 0 indicates better clustering. The obtained value 1.25, which indicates clusters are somewhat well separated.

PCA visualization shows the clusters in 2D space. Based on the PCA plot, we can observe that some clusters formed by K-Means may not be well-separated, and there may be irregularly shaped clusters.

Proposed Improvement:

Given the irregular shapes of clusters, DBSCAN(Density-Based Spatial Clustering of Applications with Noise) excels when the data contains clusters of arbitrary shapes and varying densities.