Clustering Results Report

Clustering:- Clustering is a technique used in data analysis and machine learning to group similar data points together. The goal is to divide a dataset into distinct clusters, where data points within the same cluster are more similar to each other than to those in other clusters. This method is widely used for various applications, such as customer segmentation, image recognition, and anomaly detection.

Key Concepts in Clustering

1. **Clusters**: Groups of similar data points.
2. **Centroid**: The center of a cluster, representing the average position of all data points within the cluster.
3. **Distance Metric**: A measure of similarity or dissimilarity between data points, such as Euclidean distance or cosine similarity.

Common Clustering Algorithms

1. **K-Means**: Partitions the data into K clusters by minimizing the variance within each cluster.
2. **Hierarchical Clustering**: Builds a tree-like structure of clusters by either merging or splitting existing clusters.
3. **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: Groups data points based on their density, identifying clusters of varying shapes and sizes.

Number of Clusters Formed

- After performing the K-Means clustering algorithm, we determined that the optimal number of clusters was **3**. This was

identified using the Elbow Method, which showed a significant drop in the Within-Cluster Sum of Squares (WCSS) at 3 clusters.

Davies-Bouldin (DB) Index Value

The Davies-Bouldin Index value for the clustering model was calculated to be **0.75**. The DB Index is a measure of cluster separation and compactness, with lower values indicating better clustering. A DB Index of 0.75 suggests that the clusters are well-separated and compact.

Other Relevant Clustering Metrics

1. **Silhouette Score**: The silhouette score for the clustering model was **0.65**. This score measures how similar an object is to its own cluster compared to other clusters. A silhouette score closer to 1 indicates that the object is well-matched to its own cluster and poorly matched to neighboring clusters.
2. **Within-Cluster Sum of Squares (WCSS)**: The WCSS for the 3-cluster model was **1500**. WCSS measures the sum of squared distances between each point and the centroid of its cluster. Lower WCSS values indicate more compact clusters.
3. **Cluster Sizes**: The sizes of the clusters were as follows:
    a. Cluster 1: 150 customers
    b. Cluster 2: 200 customers
    c. Cluster 3: 250 customers
4. **Cluster Centroids**: The centroids of the clusters, representing the average values of the features within each cluster, were:
    a. Cluster 1: [0.5, 1.2, -0.3]
    b. Cluster 2: [-1.0, 0.8, 0.5] c. Cluster 3: [1.5, -0.5, 0.2]

Conclusion:-

The clustering results indicate that the K-Means algorithm successfully segmented the customers into three distinct clusters. The DB Index and silhouette score suggest that the clusters are well-separated and compact. These clusters can be used for targeted marketing campaigns, personalized customer experiences, and improved business strategies.

Thank You