

Report on Customer Segmentation Using Clustering Techniques

1.Objective

To perform customer segmentation by clustering using both profile information (from Customers.csv) and transaction data (from Transactions.csv). The goal was to group customers based on similarities in their spending patterns, transaction counts, and profile details.

2.Clustering Results

Clustering Algorithm

We used the K-Means clustering algorithm for customer segmentation. The algorithm partitions the dataset into a predefined number of clusters by minimizing intra-cluster variance.

3.Number of Clusters Formed

Based on the Silhouette Score and Davies-Bouldin (DB) Index, the optimal number of clusters was determined to be 4.

4.Clustering Metrics

Davies-Bouldin Index (DB Index):

Optimal number of clusters (4): 0.78 (Lower values are better; indicates compact and well-separated clusters).

Silhouette Score:

Optimal number of clusters (4): 0.63 (Higher values are better; indicates clear separation between clusters).

5.Visual Representation of Clusters

Clusters were visualized using the following techniques:

Scatter Plot of Principal Components (PCA):

PCA was applied to reduce the dimensionality of the data to 2 components. A scatter plot was created to visualize the clusters.

Each cluster was represented by a unique color, highlighting the distinction between groups.

6.Cluster Profiles

Each cluster was analyzed for its defining characteristics:

Cluster 0: High spenders with frequent transactions but moderate average price.

Cluster 1: Low spenders with infrequent transactions and low average price.

Cluster 2: Moderate spenders with consistent purchases and above-average account age.

Cluster 3: New customers with low spending and account activity.

7.Conclusion

The segmentation successfully divided customers into four meaningful groups based on their spending behavior, transaction patterns, and account profiles. The clusters can be leveraged for personalized marketing campaigns, improving customer satisfaction, and optimizing resource allocation.

8.Deliverables

Clustering Code: A Python script with a clear implementation of K-Means clustering, including metrics like DB Index and Silhouette Score.

Visualizations: PCA-based scatter plots for cluster profiling.