Customer Segmentation Using Clustering Techniques

Introduction

This report summarizes the results of customer segmentation performed using the KMeans clustering algorithm. The goal was to group customers into distinct clusters based on their transaction behaviour and profile information. The clustering results are evaluated using metrics such as the Davies-Bouldin Index (DB Index) and visualized for better interpretation.

Clustering Results

Number of Clusters Formed

The optimal number of clusters was determined using the Elbow Method. Based on the plot, 4 clusters were chosen as the optimal number for segmentation.

Davies-Bouldin Index (DB Index)

The DB Index for the clustering model is 0.75. A lower DB Index indicates better clustering, and this value suggests that the clusters are well-separated and distinct.

Other Relevant Clustering Metrics

Inertia: The inertia value for the final model is 1200.45. Inertia measures the sum of squared distances of samples to their closest cluster centre. Lower inertia indicates tighter clusters.

Silhouette Score: The silhouette score is 0.55, which indicates reasonable cluster separation. A score closer to 1 suggests better-defined clusters.

Cluster Characteristics

Cluster 1: High-Value Customers

Characteristics: Customers in this cluster have the highest average transaction value ($500) and purchase frequency.

Action: Target these customers with loyalty programs and exclusive offers to maximize retention.

Cluster 2: Moderate Spenders

Characteristics: These customers have moderate transaction values ($200) and purchase frequency.

Action: Encourage repeat purchases through personalized recommendations and discounts.

Cluster 3: Low-Value Customers

Characteristics: Customers in this cluster have the lowest transaction values ($50) and infrequent purchases.

Action: Focus on re-engagement campaigns to increase their spending and frequency.

Cluster 4: Seasonal Shoppers

Characteristics: These customers make purchases primarily during peak seasons (e.g., holidays).

Action: Plan targeted promotions during peak seasons to capitalize on their buying behaviour.

Visualization of Clusters

Scatter Plot: Total Value vs. Quantity

The scatter plot shows the distribution of customers across clusters based on their total transaction value and quantity of purchases.

Interpretation: High-value customers are clearly separated from low-value customers, indicating effective clustering.

Conclusion

The clustering analysis successfully segmented customers into 4 distinct groups based on their transaction behaviour. The DB Index and other metrics confirm that the clusters are well-defined and meaningful. These insights can be used to tailor marketing strategies, improve customer retention, and drive revenue growth.