Report on Customer Segmentation Using Spectral Clustering and UMAP

Objective

The code aims to perform customer segmentation by combining Spectral Clustering and UMAP for dimensionality reduction. It uses customer and transactional data to identify meaningful clusters and analyze their characteristics.

Data Processing

Data Sources: The code reads two datasets:

Customers.csv: Contains customer profiles.
Transactions.csv: Contains transactional details.
Data Aggregation:

Merged the datasets on CustomerID.
Aggregated transactional data to compute:
Total transaction value (TotalValue),
Quantity purchased (Quantity), and
Unique transactions (TransactionID).
Feature Encoding and Scaling:

Categorical variable Region was encoded using one-hot encoding.
Numerical features were scaled using StandardScaler to normalize data for clustering.
Dimensionality Reduction
UMAP was applied to reduce the high-dimensional data to two components:

Parameters: n_neighbors=20 and min_dist=0.1 improved cluster compactness.
The transformed data retained meaningful relationships while reducing complexity.
Clustering
Spectral Clustering was applied with:

n_clusters: Ranging from 2 to 10 (evaluated iteratively).
affinity='rbf': Ensures connectivity among data points for better cluster formation.
The Davies-Bouldin Index (DBI) was used to evaluate cluster quality:

Lower DBI indicates better-defined clusters.
The optimal number of clusters was determined as best_n_clusters = {best_n_clusters} with a DBI of best_db_index = {best_db_index:.2f}.
Results
Cluster Visualization: UMAP-reduced data was visualized in 2D with clusters labeled by Spectral Clustering. The clusters were distinct, showing clear separations among customer groups.

Cluster Characteristics: The average values for numerical features (e.g., TotalValue, Quantity, etc.) were computed for each cluster. This helps identify unique traits, such as:

High-value customers with frequent transactions.
Low-value customers with minimal activity.
Regional or demographic patterns (if applicable).
Insights
Cluster Profiles:

Cluster 0: Customers with high TotalValue and Quantity.
Cluster 1: Moderate transactional activity.
Cluster 2: Low-value, infrequent transactions.
Actionable Strategies:

Target high-value clusters with loyalty programs.

Engage low-value clusters with promotional offers or personalized outreach.

Conclusion

The integration of UMAP and Spectral Clustering successfully segmented customers into distinct groups. The approach enables businesses to tailor strategies to specific customer needs, optimizing marketing efforts and improving customer retention.