Lookalike Model Report

Overview

The Lookalike Model identifies customers with similar purchasing behavior and spending patterns. By analyzing transaction history, product preferences, and spending habits, the model recommends the top 3 most similar customers for a given user.

Approach

1. **Data Preprocessing:**
    - Merged **`Customers.csv`**, **`Products.csv`**, and **`Transactions.csv`** to create a unified dataset.
    - Extracted key features: total spending (`TotalValue`), transaction count, average purchase price, and most purchased category.
    - Encoded categorical features (e.g., product category) for numerical processing.
2. **Feature Scaling & Transformation:**
    - Used **StandardScaler** to normalize numerical features, ensuring fair comparison.
    - Applied weighted cosine similarity and Manhattan distance-based clustering for better similarity measurement.
3. **Model Development:**
    - Implemented **K-Nearest Neighbors (KNN)** with optimized parameters (**`n_neighbors=4, metric="euclidean"`**) to find the top 3 most similar customers.
    - Assigned similarity scores using a transformed inverse distance formula to ensure meaningful ranking.
4. **Lookalike Recommendation System:**
    - For each customer, retrieved the 3 closest customers based on similarity scores.
    - Stored results in a structured format (**`Lookalike.csv`**) with **CustomerID → (LookalikeID, Similarity Score)** mapping.
    - Enabled user input to find similar customers dynamically.

Key Insights

- The model effectively clusters customers with similar spending and product preferences.
- Higher similarity scores indicate strong behavioral alignment, improving marketing strategies.
- The weighted approach reduces noise, making recommendations more precise.

This approach enhances customer understanding, enabling businesses to **boost engagement, personalize experiences, and optimize retention strategies.**