Pranav Dhobi - Lookalike Model Report

1. Introduction

In this task, we developed a Lookalike Model to recommend the top 3 similar customers for the first 20 customers (C0001 to C0020) based on their profiles and transaction histories. The model uses customer and product information to calculate similarity scores, providing personalized recommendations.

2. Approach

1. Data Preprocessing:

- Merged datasets (Customers.csv, Products.csv, and Transactions.csv) to create a unified dataset with customer, product, and transaction information.
- Created a Customer-Product Interaction Matrix, where rows represent customers, columns represent products, and values represent the quantity purchased.

2. Feature Standardization:

- Standardized the interaction matrix using StandardScaler to ensure all features contribute equally to similarity calculation.

3. Similarity Computation:

- Used Cosine Similarity to compute similarity scores between customers based on their purchase patterns and preferences.

4. Top-3 Recommendations:

- For each customer, sorted other customers by their similarity scores in descending order.
- Selected the top 3 most similar customers for each of the first 20 customers (C0001 to C0020).

4. Evaluation of Model Accuracy

1. Consistency:

- Recommendations align logically with customer profiles and transaction history.

- Customers with overlapping product categories and regions are consistently recommended.

- Example: C0001 and C0012 share similar product preferences in the 'Electronics' category.

2. Diversity:

- The top 3 recommendations for each customer are unique and not repetitive across the results.

3. Scalability:

- The model works efficiently for the first 20 customers but can be extended to larger datasets.