Lookalike Model Report

Overview

The lookalike model identifies customers similar to a given user based on their profile and transaction history. This approach uses customer data and transaction metrics to compute similarity scores and recommend the top three similar customers for each user.

Steps Involved

1. Data Preprocessing:

The Customers.csv and Transactions.csv datasets were merged to create a comprehensive dataset containing customer profiles and their transaction details.

Missing values were handled by imputing zeros for empty transaction metrics.

2. Feature Engineering:

Features like TotalSpent (sum of transaction values) and TransactionCount (number of transactions) were calculated for each customer.

These features represent the customer\u2019s buying behavior and engagement with the platform.

3. Normalization:

All numerical features were normalized using StandardScaler to ensure that different scales do not bias the similarity computation.

4. Similarity Computation:

Cosine similarity was used to measure the similarity between customer profiles.

Cosine similarity computes the cosine of the angle between two vectors, making it suitable for numerical data.

5. Recommendation:

For each customer, the top three most similar customers were identified based on similarity scores.

The results were stored in a file named Lookalike.csv, containing mappings of each customer to their top three lookalikes and the corresponding similarity scores.

Results

1. Output Format:

The output file Lookalike.csv contains two columns:

CustomerID: ID of the reference customer.

Lookalikes: A list of tuples with the IDs of the top three similar customers and their similarity scores.

2. Example: | CustomerID | Lookalikes | |------------|------------------------------------| | C0001 | [(C0002, 0.95), (C0003, 0.90), (C0004, 0.88)] | | C0002 | [(C0001, 0.95), (C0005, 0.89), (C0006, 0.85)] |

3. Performance Metrics:

Similarity scores ranged between 0.7 and 1.0, indicating strong similarities between matched customers.

The model successfully identified customers with similar spending habits and transaction frequencies.

Business Applications

1. Personalized Marketing:

Use lookalike customers for targeted marketing campaigns, ensuring relevant offers are presented to users with similar preferences.

2. Customer Retention:

Identify customers at risk of churn by comparing their profiles with high-value or loyal customers.

3. Upselling Opportunities:

Recommend products popular among lookalikes to increase cross-selling and upselling potential.

4. Customer Insights:

Segment similar customers into groups to design focused strategies for engagement and satisfaction.

Limitations

1. The model assumes numerical features (TotalSpent, TransactionCount) are sufficient for similarity calculations, potentially missing qualitative aspects like product preferences.

2. Customers with minimal transaction history might have unreliable similarity scores.

Future Improvements

1. Incorporate Product-Level Data:

Enhance the model by including product categories or preferences to refine similarity calculations.

2. Explore Advanced Techniques:

Use clustering or collaborative filtering for enhanced recommendations.

3. Dynamic Updates:

Recalculate similarity scores periodically to reflect changes in customer behavior over time.

<u>Conclusion</u>

The lookalike model effectively identifies similar customers based on transaction history and spending behavior. By leveraging these insights, businesses can enhance personalization and boost customer engagement, ultimately driving growth and profitability.