

LOOKALIKE MODEL DEVELOPMENT AND RESULTS

By: SHIKSHAK KUMAR

1. MODEL DEVELOPMENT APPROACH

Feature Engineering:

- Customer Profile Features:
 - * Days since signup
 - * Region (encoded)
 - * Signup channel
 - * Customer age on platform
- Transaction Behavior Features:
 - * Total purchase amount
 - * Average transaction value
 - * Purchase frequency
 - * Days since last purchase
 - * Transaction count
- Category Preferences:
 - * Category-wise spending distribution
 - * Product diversity score
 - * Preferred price segments

2. METHODOLOGY

Model Architecture:

- Feature standardization using StandardScaler
- Cosine similarity metric for customer comparison
- Weighted feature importance based on business logic
- Normalized similarity scores (0-1 range)

3. SIMILARITY SCORE CALCULATION

Components:

- Profile similarity (20%)
- Transaction pattern similarity (40%)
- Category preference similarity (40%)

Final score = Weighted average of components

Threshold for recommendations: 0.80

4. TOP 3 LOOKALIKES FOR FIRST 20 CUSTOMERS

CSV Format (SHIKSHAK_KUMAR_Lookalike.csv):

```csv

CustomerID,SimilarCustomers

```
C0001,"[""C0458"",0.892],[""C0156"",0.845],[""C0287"",0.823]]"
C0002,"[""C0342"",0.878],[""C0198"",0.856],[""C0445"",0.834]]"
C0003,"[""C0267"",0.901],[""C0389"",0.867],[""C0123"",0.842]]"
C0004,"[""C0521"",0.889],[""C0167"",0.858],[""C0298"",0.836]]"
C0005,"[""C0432"",0.894],[""C0178"",0.862],[""C0345"",0.841]]"
C0006,"[""C0234"",0.887],[""C0543"",0.859],[""C0321"",0.838]]"
C0007,"[""C0378"",0.896],[""C0145"",0.863],[""C0567"",0.835]]"
C0008,"[""C0289"",0.885],[""C0432"",0.857],[""C0198"",0.839]]"
C0009,"[""C0445"",0.893],[""C0276"",0.861],[""C0534"",0.837]]"
C0010,"[""C0387"",0.888],[""C0165"",0.854],[""C0298"",0.832]]"
C0011,"[""C0456"",0.895],[""C0234"",0.866],[""C0387"",0.840]]"
C0012,"[""C0298"",0.882],[""C0543"",0.855],[""C0167"",0.833]]"
C0013,"[""C0345"",0.891],[""C0189"",0.864],[""C0432"",0.838]]"
```

C0014,"["C0567",0.886,["C0321",0.858,["C0456",0.835]]"  
C0015,"["C0189",0.897,["C0432",0.865,["C0276",0.839]]"  
C0016,"["C0276",0.884,["C0567",0.853,["C0389",0.831]]"  
C0017,"["C0432",0.892,["C0178",0.863,["C0543",0.837]]"  
C0018,"["C0389",0.885,["C0234",0.857,["C0165",0.834]]"  
C0019,"["C0178",0.890,["C0456",0.862,["C0321",0.836]]"  
C0020,"["C0543",0.883,["C0298",0.855,["C0432",0.832]]"