Customer Behavior Prediction Model Using RFM and Logistic Regression

Final Project

Latifah Mohammed AL-Rasheedi
12-15-2021

Design

This project is about creating a model that can help in identifying loyal customers and predicting their next day of purchase. It would allow marketing teams to create targeted campaigns based on that information.

Data

The data used in this project is an E-Commerce dataset from Kaggle, which contains actual transactions from UK retailers.

Number of features: 8

Number of observations: 541,909

The **target** that I used is day_range, which I added to the data. The feature contains two values:

- 0 for customers who didn't purchase within 3 months.
- 1 for customers who did make a purchase within 3 months.

Algorithms

For feature engineering

- Added 3 new features to the data:
  total_amount: Quantity * unit_price
  next_purchase_date: first purchase day - last purchase date.
  day_range: 0 or 1.
- Convert categorical values to numerical using dummy.
- Split the invoice_date to: year, month, day, and hour.

For RFM method

RECENCY (R): Days since last purchase

FREQUENCY (F): Total number of purchases

MONETARY VALUE (M): Total money this customer spent

For data preprocessing

I split the data into 80% training and 20% testing.

For modeling

I used three models:

- Logistic regression and it scored 88% in accuracy.
- Random Forest Classifier and it scored 87.8% in accuracy.

- KNN classifier and it scored 75% in accuracy.

Tools

- Language: Python.
- Environment: Jupyter Notebook.
- For data processing: panda, NumPy.
- For data visualization: Seaborn, Matplotlib, Plotly.
- Algorithms and modeling: Scikit-learn, Logistic Regression model, RFM Model for customer segmentation, K-Means Clustering.

Communication

Presented the findings as presentation file.