

Team-2

Customer Segmentation Documentation

Team Members:

R. Vishnu Teja - 2011CS020319

M. Abhilash - 2011CS020210

M. Manish Yadav - 2011CS020225

K. Akhilesh - 2011CS020172

N.Aravind – 2011CS020254

Introduction:

Today many of the businesses are going online and, in this case, online marketing is becoming essential to hold customers, but during this, considering all customers as same and targeting all of them with similar marketing strategy is not very efficient way rather it's also annoys the customers by neglecting his or her individuality, so customer segmentation is becoming very popular and also became the efficient solution for this existing problem. Customer segmentation is defined as dividing company's customers on the basis of demographic (age, gender, marital status) and behavioral (types of products ordered, annual income) aspects. Since demographic characteristics does not emphasize on individuality of customer because same age groups may have different interests so behavioral aspects is a better approach for customer segmentation as its focus on individuality and we can do proper segmentation with the help of it.

Libraries used:

- pandas
- numpy
- scipy
- scikit-learn
- matplotlib
- seaborn
- nltk
- wordcloud
- jupyter notebook

Pre-processing the data:

The raw data we downloaded is complex and in a format that cannot be easily ingested by customer segmentation models. We need to do some preliminary data preparation to make this data interpretable.

The informative features in this dataset that tell us about customer buying behavior include “Quantity”, “InvoiceDate” and “UnitPrice.” Using these variables, we are going to derive a customer’s RFM profile - Recency, Frequency, Monetary Value.

RFM is commonly used in marketing to evaluate a client’s value based on their:

1. Recency: How recently have they made a purchase?
2. Frequency: How often have they bought something?
3. Monetary Value: How much money do they spend on average when making purchases?

With the variables in this e-commerce transaction dataset, we will calculate each customer’s recency, frequency, and monetary value. These RFM values will then be used to build the segmentation model.

Algorithms Used:

- SVM
- Logistic Regression
- Decision Tree
- Random Forest
- Ada Boost Classifier
- Gradient Boosting Classifier
- **Evaluation Metrics:**

Precision:

Precision is one indicator of a machine learning model's performance – the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives).

Results:

