

Thesis Project Portfolio

Customer Segmentation using RFM Analysis and K-Means Clustering to enhance Marketing Strategies

(Technical Report)

Analysis of Algorithmic Bias in Customer Segmentation Models

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Harshil Anishkumar Patel

Spring, 2023

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Customer Segmentation using RFM Analysis and K-Means Clustering to enhance Marketing Strategies

Analysis of Algorithmic Bias in Customer Segmentation Models

Prospectus

Sociotechnical Synthesis

My technical and STS research projects are connected by the idea of customer segmentation using machine learning algorithms. Customer segmentation refers to the process of dividing a customer base into groups based on common characteristics, such as demographics, behaviors, or preferences. Machine learning algorithms can help businesses analyze large amounts of customer data and gain insights into customer behavior and preferences, enabling them to develop targeted marketing strategies and improve performance. However, there are potential ethical concerns related to the use of machine learning algorithms in customer segmentation, as these algorithms may introduce biases that could have negative consequences for some customer segments.

In my technical report, I aim to demonstrate a customer segmenting approach using the k-means clustering algorithm and RFM analysis model to create clusters of customers based on their purchasing behavior. The project's objective is to help businesses target customers effectively, resulting in improved performance and customer services. By utilizing the RFM model and k-means clustering, we can group customers into four distinct clusters based on their recency, frequency, and monetary value. This segmentation allows businesses to identify high-value and low-value customer segments and those that require targeted marketing efforts. The project methodology comprises five stages: business understanding, data understanding, data preparation, modeling, and evaluation. The results show that the k-means clustering algorithm, coupled with the RFM model, is effective in clustering customers based on their purchasing behavior. The insights from the model enable businesses to analyze their customers, allowing them to develop targeted marketing strategies that enhance customer retention and acquisition. Future work of this

project involves incorporating additional data sources such as, tenure, enhancing the web model's features, and conducting additional testing and evaluation to improve the clustering accuracy.

In my STS research project, I explore the impact of AI-powered customer segmentation models on businesses and consumers. I analyze various approaches to customer segmentation models that were meant to define target markets and precisely identify customer segments, but biases affected the system along the way. I also evaluate the limitations of these models within the lens of the Actor-Network Theory and Social Construction of Technology. The study highlights the need for companies to prioritize diversity and inclusivity in their data collection and analysis processes to avoid biases being introduced into these customer segmentation models, which can have negative consequences. Ultimately, the responsible integration of AI in customer segmentation models can help companies build trust with their customers, create personalized experiences, and drive growth in the increasingly competitive business landscape.

By exploring these two projects simultaneously, I gained an understanding of why and how machine learning algorithms can be used to create customer segments and the potential ethical concerns associated with their use. While the technical report focused on developing an effective customer clustering approach using k-means clustering and the RFM model, the STS research paper evaluated the potential limitations and ethical considerations related to the use of AI technology in customer segmentation models. By examining these two projects side by side, I realized the importance of considering the ethical implications of machine learning algorithms in developing effective marketing strategies. The responsible use of AI in customer segmentation models can lead to better customer experiences and enhanced business performance while avoiding potential ethical concerns.