Customer Segmentation in Marketing using Machine Learning

Kritant Kumar[1], Balajee Nihal[2] , Ritesh Kumar[3], Priyanshu Kumar [4]

[1,2,3,4] IIMT College of Engineering, AKTU, Greater Noida, India.
[1] kritantvbnm@gmail.com
[2] balajee.nihall@gmail.com
[3] rchaudhari2314@gmail.com
[4] priyanshu07264@gmail.com

Abstract- The emergence of many competitors and entrepreneurs has caused a lot of tension among competing businesses to find new buyers and keep the old ones. As a result of the predecessor, the need for exceptional customer service becomes appropriate regardless of the size of the business. Furthermore, the ability of any business to understand the needs of each of its customers will provide greater customer support in providing targeted customer services and developing customized customer service plans. This understanding is possible through structured customer service. Each segment has customers who share the same market features. Big data ideas and machine learning have promoted greater acceptance of automated customer segmentation approaches in favor of traditional market analytics that often do not work when the customer base is very large. In this paper, the k-means clustering algorithm is used for this purpose. The Sklearn liabrary was developed for the k-Means algorithm and the program is trained using a 100-pattern two-factor dataset derived from the retail trade. Characteristics of average number of customer purchases and average number of monthly customers.

Keywords: Customer Segmentation, Algorithms, Machine Learning (ML), Artificial Language (AI), Detailed Analysis, Survey, Github, Marketing

1 Introduction

Over the years, increased competition among businesses and the availability of large-scale historical data has resulted in widespread use of data mining techniques to find critical and strategic information that is hidden in organizations' information. Data mining is the process of extracting logical information from a dataset and presenting it in a human-accessible manner for decision support. Data mining techniques distinguish fields such as statistics, artificial intelligence, machine learning, and data systems. Data mining applications include, but are not limited to bioinformatics, weather forecasting, fraud detection, financial analysis and customer segmentation. The key to this paper is to identify customer segments in a commercial business using the data mining method. Customer segmentation is a group of business customer base called customer segment such that each customer segment has customers who share the same market characteristics. These differences are based on factors that directly or indirectly affect the market or business such as product preferences or expectations, location, behavior and so on. The importance of customer segmentation includes, inter alia, the ability of a business to customize market plans that would be appropriate for each segment of its customers. Support for business decisions based on risky environments such as credit relationships with its customers; Identify products related to individual components and how to manage demand and supply power; Interdependence and interaction between consumers, between products, or between customers and products are revealed, which the business may not be aware of; The ability to predict customer declines, and which customers are likely to have problems and raise other market research questions and provide clues to find solutions. Buried in a database of integrated data proved to be effective for detecting subtle but subtle patterns or relationships. This mode of learning is classified under supervised learning. Integration algorithms include the KMeans algorithm, K-nearest algorithm, sorting map (SOM), and more. These algorithms, without prior knowledge of the data, are able to identify groups in them by repeatedly comparing input patterns, as long as static aptitude in training examples is achieved based on subject matter or process. Each set has data points that have very close similarities but differ greatly from the data points of other groups. Integration has great applications in pattern recognition, image analysis, and bioinformatics and so on. In this paper the k-means clustering algorithm was implemented in the customer segment. The scalar library of the K-Means algorithm was developed, and training was started using a standard silhouette -score with two feature sets of 100 training patterns found in the retail trade. After several indications, four stable intervals or customer segments were identified. Two factors are considered in combination with the number of items a customer purchases per month and the average number of customers per month.

2 Progression of Customer Segmentation

Over the years, the commercial world has become more competitive, as organizations such as these have to meet the needs and desires of their customers, attract new customers, and thus improve their businesses. The task of identifying and meeting the needs and requirements of every customer in the business is very difficult. This is because customers can vary according to their needs, wants, demographics, size, taste and taste, features etc. As it is, it is a bad practice to treat all customers equally in business. This challenge has adopted the concept of customer segmentation or market segmentation, where consumers are divided into subgroups or segments, where members of each subcategory exhibit similar market behaviors or characteristics. Accordingly, customer segmentation is the process of dividing the market into indigenous groups.

2.1 Direct Survey

Recently, Big Data research has gained momentum. Defines big data - a term that describes a large number of formal and informal data, which cannot be analyzed using traditional methods and algorithms. Companies include billions of data about their customers, suppliers, and operations, and millions of internally connected sensors are sent to the real world on devices such as mobile phones and cars, sensing, manufacturing and communications data. Ability to improve forecasting, save money, increase efficiency and improve various areas such as traffic control, weather forecasting, disaster prevention, finance, fraud control, business transactions, national security, education and healthcare. Big data is mainly seen in three Vs: volume, variability, and speed. Other 2Vs are available - authenticity and price, thus making it 5V.

2.2 Online Survey

Clustering is the process of grouping information into a dataset based on some commonalities. There are several algorithms, which can be applied to datasets based on the provided condition. However, no universal clustering algorithm exists, hence it becomes important to choose the appropriate clustering techniques. In this paper, we have implemented three clustering algorithms using the Python scalar library. E. K-mein K-means that an algorithm is one of the most popular classification algorithms. This clustering algorithm relies on centro, where each data point is placed in one of the overlapping ones, which is pre-sorted in the K-algorithm. Clusters are created that correspond to hidden patterns in the data that provide the necessary information to help decide execution. process. There are many ways to make assembling K-means, we will use the elbow method.

2.3 Advent of Digital Survey

There are many ways to partition, which vary in severity, data requirements, and purpose. The following are some of the most commonly used methods, but this is not an incomplete list. There are papers that discuss artificial neural networks, particle determination and complex types of ensemble, but are not included due to limited exposure. In future articles, I may go into some of these options, but for now, these general methods should suffice. Each subsequent section of this article will include a basic description of the method, as well as a code example for the method used. If you do not have the expertise, well, just skip the code and you have to get a good handle on each of the 4 sub-sections included in this article.

2.4 Artificial Intelligence Era

: The role of market segmentation in shaping pricing strategies for new products is critical. This study highlights the significance of tailoring pricing decisions in meeting the unique needs, preferences, and price sensitivities of different consumer segments in order to maximize profitability and market penetration. Understanding customer behavior and preferences in transitioning from high to low prices during new product introductions is crucial. The responses of various customer groups to price changes are essential in influencing product innovation. The model developed in this study serves two purposes: to determine the optimal time to switch from high to low prices, and to determine the optimal price discount when switching from high to low prices for different customer segments. Segmented marketing results in larger profits due to increased sales to loyal customers. However, deal-prone customers may purchase less when segmented.

2.4.1 Why ML?

In the realm of consumer electronics, companies such as Apple have mastered the art of introducing new products at premium prices. For instance, when a new iPhone model hits the market, early adopters and brand enthusiasts eagerly purchase it at a high price point. These “loyal” customers value the latest features and are willing to pay a premium for the innovative technology. This initial pricing strategy not only caters to brand loyalists but also serves as a revenue-maximizing tactic. However, as time progresses and competition intensifies, the price of the new iPhone gradually declines to attract a broader customer base, including those who prioritize affordability.

3 ML methods for Customer Segmentation

Companies need to understand the diverse needs, preferences, and price sensitivities of different consumers in order to tailor their pricing, product quality, and marketing efforts to specific customer groups. This is crucial for maximizing profitability and market penetration while transitioning from high to low prices. Market segmentation, which involves categorizing potential buyers into groups based on common needs and responses to marketing initiatives, serves as a guiding framework for pricing decisions tailored to the unique characteristics of each segment (group). This ultimately influences the success of new product launches. By segmenting the market, marketers can target specific customer groups with pricing, product quality, and marketing efforts customized to their preferences and behaviors. This enables them to make informed decisions about when and how to effectively apply price discounts to different segmented customer groups. Therefore, market segmentation plays a crucial role in the transition from high to low pricing during new product introductions.

3.1 K-Means Clustering

When a new EV is introduced to the market (see for example the case of Tesla, it often comes with a relatively high price. This high price is typically set to target loyal customers or early adopters who prioritize environmentally friendly features and are willing to pay a premium for the latest technology and clean energy benefits. The high initial price is justified by the inclusion of exclusive features, superior performance, longer- lasting batteries, and cutting-edge technology. This attracts customers who value these attributes. During the initial phase, the manufacturer focuses on selling to environmentally conscious and brand-loyal customers who are willing to pay the high price for the premium product. This phase allows the manufacturer to establish a presence in the market and recover some of the initial development costs. As time progresses, and the EV matures in the market, the manufacturer may consider reducing the price to attract a broader customer base, including deal-prone customers who are more price-sensitive. The price reduction strategy may involve offering discounts, incentives, or introducing more affordable EV models within the brand’s portfolio. With lower prices and more accessible options, the EV manufacturer can tap into a larger segment of the market. This includes customers who are concerned about the environment but prioritize the price more highly. The goal is to boost the volume of sales and to increase market share by making EVs more affordable and appealing to a wider audience. The timing of the price switch from high to low prices is influenced by the product’s lifecycle. This switch typically occurs when the product has reached a certain level of market saturation and competition has intensified.

3.2 DBSCAN

Companies need to understand the diverse needs, preferences, and price sensitivities of different consumers in order to tailor their pricing, product quality, and marketing efforts to specific customer groups. This is crucial for maximizing profitability and market penetration while transitioning from high to low prices. Market segmentation, which involves categorizing potential buyers into groups based on common needs and responses to marketing initiatives, serves as a guiding framework for pricing decisions tailored to the unique characteristics of each segment . This ultimately influences the success of new product launches. By segmenting the market, marketers can target specific customer groups with pricing, product quality, and marketing efforts customized to their preferences and behaviors. This enables them to make informed decisions about when and how to effectively apply price discounts to different segmented customer groups. Therefore, market segmentation plays a crucial role in the transition from high to low pricing during new product introductions.

3.3 Hierarchical Clustering

To elucidate the concept of market segmentation, consider for example a multinational fast-food chain like McDonald’s. The fast-food industry employs market segmentation by identifying different customer segments based on behavior, demographic criteria, and cultural attributes. For instance, they may target families with young children looking for kid-friendly meal options or Generation Y students who are considered to be the key market segment in the fast-food industry, with significant purchasing power due to their lifestyle and eating habits. At the same time, they might also target health-conscious consumers seeking specific menu choices. These distinct customer segments have varying needs and preferences, demanding tailored marketing strategies, menu offerings, and pricing structures.

4 Successful ML Case Studies in Customer Segmentation

Market segmentation within the clothing retail industry groups consumers based on their consumer behavior, motivations, and psychographics. A clothing retailer may segment its market based on age, gender, fashion involvement, and lifestyle. They may offer different styles and sizes of clothing to appeal to different segments, such as athletic wear for younger customers, business attire for adult customers or luxury fashion to appeal to various customer preferences, such as ready-to-wear, haute couture, or limited-edition pieces. Clothing retailers should align their price transitions during new product launches with their market segmentation strategies. By understanding the preferences and behaviors of each segment, retailers can set initial prices that resonate with their target customers and then adjust pricing strategies over time to optimize sales and profitability while maintaining customer loyalty and satisfaction.

4.1 Virtual Analysis

Market segmentation is a crucial strategy for a global beverage company that introduces a new line of fruit- flavored soft drinks. Instead of applying a one-size-fits-all approach, the company conducts market research to identify different consumer segments based on factors such as age, taste preferences, and dietary choices. This segmentation reveals various clusters of consumers, including health-conscious millennials who prioritize low- calorie options, families seeking natural and sugar-free alternatives, and adventureseekers looking for bold and exotic flavors. The Global Beverage Company can use a dynamic pricing approach during new product launches to cater to different market segments. Pricing transitions should align with the perceived value of the product, the preferences of each segment, and the product's market positioning. By adapting prices over time and offering promotions strategically, the company can maximize sales and market penetration within each targeted consumer segment while maintaining customer loyalty and satisfaction.

4.2 Remote Analysis

Our paper delves into the intricacies of market segmentation in the context of price transitions during new product launches. We will explore the interplay between brandloyal customers and deal-prone customers, examining how their distinct preferences and price sensitivities influence pricing strategies and ultimately impact profitability. Through theoretical modeling, we aim to shed light on the optimal timing of price switches and on the most effective price discount strategies for segmented customer groups. In doing so, we believe we are contributing valuable insights to the ongoing dialogue surrounding pricing strategies, customer segmentation, and profit optimization in the ever-evolving world of marketing. Market segmentation and pricing have long been recognized as essential components of successful marketing strategies. The academic literature on market segmentation and pricing is rich and multifaceted and reflects the profound impact of these concepts on marketing strategies and business profitability. Scholars and researchers have delved into various aspects of segmentation, including geographic, demographic, psychographic, and behavioral segmentation criteria.

4.3 Machine Learning

Marketing strategies that promote special offers—such as highlight discounts, promotions, marketing material cost savings, targeted advertising, and email campaigns—employ dynamic pricing strategies to adjust prices based on demand and competitor pricing. The characterization of deal-prone consumers has traditionally been related to price sensitivity. Price sensitivity refers to how individuals perceive and respond to changes in prices of ML. Most marketing literature assumes a direct relationship between price sensitivity and deal-proneness. Many economists and marketing experts, such as Özkan and Evrim and Joshph et al., distinguish between loyal customers with low price sensitivity, who demand a high-quality, longer-lasting, sophisticated product, and deal-prone consumers, who are sensitive to price changes and prefer a basic, shorter-lasting, lower quality product. Traditionally, a strictly negative relationship is assumed between being deal-prone and being brand-loyal. Market segmentation and pricing strategies are intertwined elements of marketing, influencing each other's outcomes. The literature emphasizes that understanding consumer heterogeneity through segmentation is crucial for effective pricing decisions, as it allows businesses to align pricing strategies with the unique preferences and behaviors of each customer segment.

5 Challenges and Future Directions in ML for Customer Segmenatation

5.1 Challenges

No technology however easy life it makes for the human being is not foolproof, some uncertainties ought to be there and there would be one or the other challenges which could occur while trying to make full use of the technology. Explainable artificial intelligence methods have many types of challenges associated with them which need to be addressed to be able to use the methods to their proper fructification.

5.1.1 Complex Data

Someone who is suffering from Diabetes and undergoing treatment requires a lot of care to be taken if the situation is on the end of the scale and the problem increases even more if it is a situation when anytime anything can happen. For this type of case, it becomes a necessity that almost all of the factors which could influence the health of that person are monitored and taken care of. Now if almost all the factors are taken care of then it would also lead to the generation of a lot of data which if taken with each passing time could lead to a lot of complexity. The various readings involved could be the readings of glucose levels, the meal which is consumed by the patient, the amount of physical activity in which the patient is involved and if applicable the doses of insulin the patient is provided with.

5.1.2 Model Comprehensibility

The innovation implementation generates costs in order to guarantee less pollution. From a short-term perspective, it increases expenses. In making green consumption decisions, consumers are faced with a social dilemma: either they can behave in an environmentally friendly manner and contribute to society by reducing pollution or they can try to maximize their own gains.

5.1.3 Processing in Real Time

A mutual consensus claimed by Policarpo and Aguiar is that customers who prefer environmentally friendly products are considered to be of a higher social status and, in our particular analysis, share the characteristics of altruists who are willing to pay a higher price for a product that sometimes underperforms compared to its conventional counterparts but is better for society as a whole. The market that is thus created includes two segments of customers: those who desire a cleaner environment and are willing and able to bear the costs of these innovations ("loyal customers"), and those who cannot afford to place a cleaner environment as a higher priority than item cost in their consumption behavior ("deal-prone customers").

5.1.4 Integration Problem

When EVs were first introduced to the market at a very high retail price, they were sold to loyal customers only. At a certain point after their entry, prices dropped dramatically, and EVs began to be sold to deal-prone customers who benefitted from the price drop until the cars were taken off the market. In our current paper, we assume that the demand of the deal-prone customers is generated and affected positively by the duration of product use and satisfaction by loyal customers (positive externalities effect of loyal customers on deal prone customers).

5.1.5 Ethical and Regulation Concerns

In the contemporary marketplace, a monopoly wields substantial influence, particularly in scenarios where it caters to multiple consumer segments characterized by differing levels of loyalty towards its products. This study focuses on two distinct equilibrium models aimed at optimizing the profit-maximizing strategies adopted by the monopoly in response to varying consumer demands. The first consumer segment, referred to as type H consumers, exhibits strong brand loyalty. They are not only willing to pay a premium price for the product but also demonstrate a proactive interest in acquiring it as soon as it becomes available. Conversely, type L consumers represent a group less committed to brand loyalty. They tend to seek deals and discounts, thus indicating a willingness to pay a lower price for the product. Their interest in the product is highly dependent on its perceived value in the marketplace.

5.1.6 Resource and Cost Constraints

Our study's first model analyzes a simplified one-period framework involving price discrimination. In this context, the monopoly offers the product to both consumer types at the beginning of the period, but with distinct pricing strategies. Type H loyal consumers are charged a premium price, while type L deal-prone customers are offered the product at a lower price point. As a result, this model involves the analysis of two decision variables, which represent the quantities that each consumer type will purchase.

5.2 Future Directions

The second model adopts a segmentation approach, introducing temporal separation between the two consumer types. During the initial period, only type H consumers purchase the higher-quality product. In the subsequent period, type L consumers enter the market and choose the basic product. Importantly, the willingness of type L consumers to pay for the product during the second period is influenced by a positive external effect, specifically represented by the duration for which type H consumers have been using the product. In this second model, the monopoly is confronted with three decision variables, encapsulating the quantities that both consumer types will purchase, along with the pivotal "switching point" that marks the transition from the purchase and use of the high- quality product by type H consumers to the purchase and use of the basic product by type L consumers. The subsequent sections of this study will provide a detailed exposition of these two models, beginning with an exploration of the first case, followed by an in-depth analysis of the second case.

5.2.1 Digital Devices

Our examination commences with the simple discrimination model. Within this framework, a monopoly confronts two linear demand curves, each associated with specific consumer categories: loyal customers and deal-prone customers. Consumers engage in concurrent product acquisition, yet they demonstrate disparate price preferences. Loyal consumers evince a predisposition to accept higher pricing, whereas deal-prone consumers exhibit a proclivity for lower price structures.

5.2.2 Personalization Facility

In this section, we will analyze a segmentation model featuring linear demand curves for the two distinct customer categories mentioned earlier. Initially, only type H consumers purchase the higher-quality product, while in the subsequent period, type L consumers enter the market and opt for the basic product. These demand curves delineate the consumption patterns exhibited by diverse customer groups, each consuming the product during distinct time intervals (as opposed to Case 1 in which consumption occurs at the same time).

5.2.3 Feedback in Real Time

ns and propositions established within the theoretical model. Upon scrutinizing the quantities sold, distinct trends emerge. Notably, in the discrimination case, the sales volume to loyal customers surpasses the volume observed in the segmentation case. Conversely, the segmentation scenario records higher sales to deal-prone customers compared with the discrimination case. Notably, as the value of parameter H increases, both cases witness an expansion in total quantities sold, gradually drawing closer to each other. Turning our attention to price dynamics, loyal customers encounter higher prices in the segmentation scenario than in the discrimination case. Similarly, deal-prone customers face elevated prices under segmentation, attributed to the influence of parameter α, which exceeds γ. Additionally, the price difference, in both cases, demonstrate a different picture.

5.2.4 Need for Standardization

In the case of segmentation, the interdependency is revealed. The initial demand is induced by loyal customers' independent demand of high price values. The length of time of product consumption by loyal customers affects deal-prone customers, and the optimization process takes into account the structure of both customer groups and the combination of each group's length of purchase and use time, which are influenced by the intertemporal (positive) effects between loyal and deal-prone customers. This effect does not occur in the regular price discrimination scenario.

5.2.5 R&D and Trials

In today's fast-paced business world, the needs and wants of customers are constantly shifting, making it a vital task for businesses to stay up-to-date and effectively connect with a diverse range of clients. The concept of customer segmentation, which involves organizing customers into distinct categories based on common qualities, has become a crucial strategy for tackling this challenge. From marketing tactics tailored to specific groups to improving overall customer satisfaction, segmentation has proven to be a valuable tool in the ever-increasing world of data-driven business. As companies navigate through the overwhelming amount of information on customer behavior, segmentation has emerged as a strategic approach to interpreting and utilizing this valuable data. In the current climate, a one-size-fits-all strategy for customer engagement is no longer effective. Nowadays, customers demand personalized interactions, tailored content, and products that speak to their unique needs. Therefore, businesses must pivot and utilize advanced data analysis techniques to deconstruct their customer demographics and identify what sets each segment apart. This not only streamlines marketing efforts, but also allows companies to adapt to evolving market trends, foster stronger connections with customers, and ultimately gain a competitive edge. This paper seeks to enhance the ongoing discourse on customer segmentation by presenting cutting-edge techniques and methodologies that harness advanced analytics, machine learning, and data-driven strategies. By examining demographics, behaviors, and transaction records, we delve into the complexities of segmentation. Through our research, we aim to tackle the obstacles and potential benefits of customer segmentation, providing a holistic understanding of how this practice can enable businesses to better cater to their customers and excel in today's data-driven landscape.

6 Conclusion

Our study introduces a cutting-edge method for customer segmentation, merging the RFM model with K-Means clustering to elevate marketing initiatives and tailor experiences for customers. Our findings demonstrate the success of this technique in categorizing customers into unique clusters according to their recent transactions, level of engagement, and monetary contributions. By utilizing the K-Means algorithm, customers were efficiently grouped revealing crucial insights into their behaviors and preferences. The practical application of our research has notably bolstered marketing efficiency and customer interaction, by providing personalized experiences and delivering more focused campaigns. As businesses increasingly seek to harness the power of data-driven strategies, the RFM and K-Means model emerges as a valuable tool for gaining a comprehensive understanding of customer behavior. This research contributes to the growing body of knowledge in customer segmentation and encourages further exploration of innovative data-driven techniques, which will play a crucial role in shaping the future of personalized marketing and customer-centric business practices.

Our research opens the door to future studies on enhancing and expanding customer segmentation techniques. To continue pushing the boundaries of this field, scholars can delve into incorporating supplemental data sources, utilizing machine learning methods, and implementing personalization tactics. Furthermore, the adaptability of our approach to various industries and larger datasets presents exciting opportunities for further investigation. The fusion of RFM-based customer segmentation and K-Means clustering offers a powerful tool for gaining deeper insights into customer behavior. It will empower data-driven decision-making for businesses and elevate customer experiences.

References

[1]. Miguel Alves et al. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases.
[2]. Prof. Nikhil Patankar et al. (2021). Customer Segmentation using Machine Learning.
[3]. Chu Chain Henry Chan et al. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer.
[4]. Swesh Raj Regmi et al. (2022). Customer Market Segmentation using Machine Learning Algorithm. [5]. Tchai Tavor et al. (2023). Customer Segmentation as a revenue generator for profit purposes.
[6]. Ishita Shah et al. (2024). Customer Segmentation.
[7]. Yash Parab et al. (2023). Customer Segmentation using Machine Learning: A Comprehensive Research Study.
[8]. Pilli Sri Durga et al. (2023). Customer Segmentation Analysis for improving sales using clustering. [9]. Jianhong Luo et al. (2022). Exploration of Spa Leisure Consumption Sentiment towards Different Holidays and Different Cities through Online Reviews: Implication for Customer Segmentation. [10]. Aman Banduni et al. (2019). Customer Segmentation using Machine Learning.