International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

CUSTOMER SEGMENTATION USING MACHINE LEARNING

Kiran D[1], Ashwini C[2]

[1]*Kiran D, Student, Department of MCA, UBDTCE, Davangere*
[2]*Ashwini C, Assistant Professor, Department of MCA, UBDTCE, Davangere*

Abstract - Businesses need to identify and segment their consumer base in order to effectively customize their strategies and improve customer satisfaction in the highly competitive market landscape of today. The goal of this study is to employ machine learning techniques to create a strong consumer segmentation model that will classify customers according to their demographics, behaviors, and purchase histories. Through the use of multiple clustering methods, including K-means, DBSCAN, and Hierarchical Clustering, the model seeks to find unique customer segments with shared attributes. To accomplish optimal segmentation, the segmentation process entails three steps: feature selection to identify the most significant features, model training, and data preprocessing to manage missing values and outliers. In-depth segment analysis is also included in the report to offer practical insights for better client retention tactics, tailored recommendations, and focused marketing efforts. The results of the study illustrate how machine learning may be used to find hidden patterns in consumer data, giving organizations the ability to make data-driven decisions. Organizations may improve their marketing efforts, allocate resources more efficiently, and eventually increase customer engagement and profitability by putting this customer segmentation strategy into practice.

Key Words: K-Means Clustering, Hierarchical Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Cluster Plotting, Heatmaps, customer relationship management (CRM) system.

1. INTRODUCTION

It's more important than ever to understand customer behavior and preferences in today's fiercely competitive business market. Massive volumes of data from several touchpoints, such as online interactions, social media engagements, and sales transactions, are flooded into companies. Strategic advantages can be gained from the efficient analysis and segmentation of this data. client segmentation is the process of breaking down a client base into discrete groups based on shared attributes. It is a basic technique that helps companies customize their marketing plans, improve customer experiences, and allocate resources as efficiently as possible.

Traditional methods for client segmentation frequently depended on basic factors like geography or demographics. These techniques are helpful, but they can't fully capture the complexity of consumer behavior. This method has been completely transformed by developments in machine learning and data analytics, opening the door to more complex and dynamic segmentation strategies. Complex datasets can be analyzed by machine learning algorithms, which can reveal hidden links and patterns that traditional methods can miss. As a result, consumer segmentation become more precise and useful, improving business outcomes.

The goal of this work is to create a strong customer segmentation model by utilizing machine learning techniques. We want to find unique consumer segments based on a broad collection of variables by employing clustering algorithms like K-means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and Hierarchical Clustering. These features give a comprehensive picture of each consumer group by incorporating demographic information, purchase patterns, and engagement indicators.

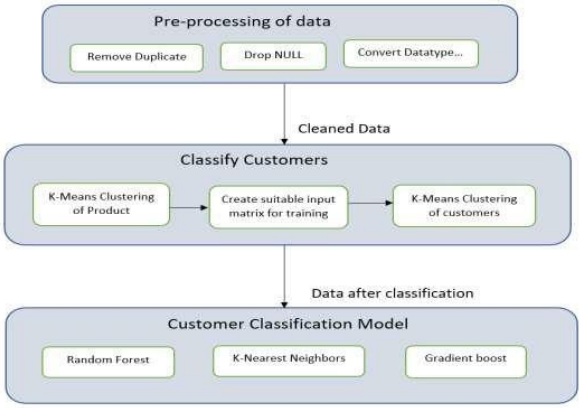


Figure 1: System architecture for customer segmentation

The methodology consists of multiple crucial processes, including feature selection to concentrate on the most pertinent qualities, data preparation to manage missing values and guarantee consistency, and data gathering from multiple sources. After that, we do exploratory data analysis (EDA) in order to obtain preliminary understanding of the data and provide guidance for the segmentation procedure. Following the application of the selected clustering algorithms, the consumers are divided into discrete segments, which are assessed and verified through the use of metrics like the Davies-Bouldin Index and Silhouette Score.

We characterize each group after the segments have been determined in order to comprehend their distinct traits and behaviors. Businesses can create personalized client encounters, more effective marketing efforts, and higher levels of customer satisfaction because to this profiling. For example, loyalty programs can target high-value groups, and re-engagement techniques can target declining engagement segments.

Significant advantages are expected from the application of this consumer segmentation methodology. Personalized services help businesses increase sales, increase client retention rates, and spend money on marketing more wisely. Additionally, by using the insights gathered from this model, businesses may improve customer service, inventory control, and product

creation in a way that is more focused on the needs of their customers.
In order to sum up, this research demonstrates how machine learning can revolutionize client segmentation. Businesses may better understand their customers, build stronger bonds with them, and promote sustainable growth by eschewing conventional approaches and utilizing advanced analytics. The approach, analysis, and outcomes of our customer segmentation model will be covered in full in the parts that follow, along with some useful applications and advantages.

2. LITERATURE SURVEY

1. T. Kansal, S. Bahuguna 2018, this paper describe Using K-means clustering for customer segmentation, the main goal is to divide the customer base into discrete groups according to shared characteristics including demographics, engagement metrics, and purchasing patterns. Businesses hope to find patterns and traits within each group by dividing their consumer base into various clusters, which enables more specialized and individualized marketing approaches. In order to better meet the demands of various client categories, segmentation aids in identifying customer preferences, raising customer happiness, and optimizing marketing initiatives.

2. Nurma Sari, Juni & Nugroho 2016, this paper describe the Customer segmentation in e-commerce aims to divide consumers into discrete groups according to their traits and habits. This makes it possible for companies to customize their offers and marketing plans to each group's unique requirements and preferences, which boosts client happiness and boosts revenue.

3. T. Kanungo, d. M. Mount 2002, the paper describe By streamlining the clustering process, an effective K-means clustering algorithm seeks to enhance the accuracy and performance of the conventional K-means approach. This entails improving the algorithm to minimize cluster assignment errors, lower computational complexity, and handle enormous datasets more skillfully. The algorithm hopes to produce quicker and more accurate clustering results by attaining these enhancements, which will increase its applicability in real-world scenarios.

4. (V.Vijiles1, 2021) This paper's main goal is to use machine learning—more specifically, the K-Means clustering algorithm—to categorize clients in an e-commerce context according to their qualities and purchasing behaviors. The article tries to uncover hidden trends in consumer data by segmenting customers into separate groups, such as High, Medium, and Low, based on their purchasing patterns, buying frequency, and recent involvement. By better targeting and attracting potential clients, e-commerce companies can outperform their rivals in the market by using this segmentation to better tailor their goods and services.

5. Chinedu Pascal, Simeon Ozuomba 2015, This paper's primary goal is to effectively segment clients according to their attributes and purchase behaviors by using the K-Means clustering technique. The study's goal is to find patterns and insights that can be applied to better target marketing campaigns and customer services by grouping customers into unique categories using K-Means. By giving more relevant and tailored experiences for various client categories, this strategy aims to increase focused consumer interaction, boost service offerings, and eventually spur corporate growth.

6. Bhade, Kalyani, et al 2018, This paper's primary goal is to offer a methodical strategy to buyer targeting and customer segmentation in order to maximize revenues. The study intends to create tailored marketing strategies and service offers by evaluating customer data to discover distinct segments and comprehending their distinctive traits and purchase behaviors. By coordinating company operations with the unique requirements and preferences of various client segments, this strategy aims to maximize revenue through raising customer happiness, boosting sales, and optimizing resource usage.

7. Pascal, C., Ozuomba, 2017, This paper's primary goal is to effectively segment customers for focused customer services by using the K-Means clustering technique. The study attempts to identify discrete consumer groups and adjust services and marketing efforts in accordance with the results by using K-Means to categorize customers based on their behavior and attributes. Through more accurate targeting and targeted customer interactions, this segmentation method aims to improve customer engagement effectiveness, service personalization, and eventually lead to improved business outcomes.

8. Tripathi, S., A. Bhardwaj 2018, This paper's main goal is to investigate and assess several clustering techniques for customer segmentation in order to determine the best strategies for breaking up customers into relevant groups. The goal of the study is to examine how well several techniques—K-Means, Hierarchical Clustering, DBSCAN, Mean Shift, Gaussian Mixture Models, and Spectral Clustering—perform in consumer segmentation based on traits and behaviors. The objective is to supply information on which clustering technique optimizes consumer segmentation and improves targeted marketing tactics by offering the optimal balance of accuracy, efficiency, and interpretability.

9. .Deng, Yulin, and Qianying Gao 2018, This study's primary goal is to improve e-commerce client segmentation through the use of an upgraded K-Means algorithm. In order to more effectively categorize clients based on their actions and traits, the classic K-Means approach is being improved. The goal of this update is to overcome the drawbacks of the original K-Means algorithm, including its sensitivity to the initial placement of the centroid and the requirement to predefine the number of clusters. The study aims to achieve more relevant

and accurate client segments through upgrades, which can result in more effective marketing tactics and better corporate decision-making in the e-commerce industry.

10. Loh, Woong-Kee 2014, This survey's primary goal is to present a thorough analysis of density-based clustering algorithms, emphasizing their underlying theories, methods, and uses. The purpose of the survey is to investigate and assess the merits, drawbacks, and applicability of a range of density-based clustering techniques, including Mean Shift, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and OPTICS (Ordering Points To Identify the Clustering Structure). The survey aims to provide insights into the efficiency of these algorithms in finding clusters of different forms and densities, managing noise, and addressing real-world data processing difficulties by providing an overview of recent research and developments in density-based clustering.

3. METHODOLOGY

3.1 Data Collection

During the data collecting process, a variety of sources of consumer information are gathered to guarantee a comprehensive understanding of the behaviors and interests of the customers. Important sources include transaction records, which provide insights into purchasing behavior like frequency and average transaction value; web analytics, which tracks engagement metrics like website visits and interaction history; and customer surveys, which collect demographic data like age, gender, and income. By combining data from several many sources, we are able to produce a reliable dataset that has all the relevant attributes needed for effective customer segmentation.

3.2 Data Preprocessing

The first stage of preparing data is data cleaning. In this case, data mistakes and inconsistencies are corrected, and missing values are handled through imputation or removal. After that, categorical variables are encoded using techniques like label encoding or one-hot encoding, and data transformation normalizes or standardizes numerical features to ensure their comparability. Last but not least, feature selection is the process of employing statistical methods or domain knowledge to determine which features are most relevant for segmentation. Any redundant or superfluous features are then removed in order to improve the accuracy and effectiveness of the model.

3.3 Segmentation Techniques

Various clustering algorithms are employed to segment the customer base. K-means clustering separates customers into discrete groups based on feature similarity; the optimal number of clusters is determined using methods such as the Elbow Method or Silhouette Analysis. Density-Based Spatial Clustering of Applications with Noise, or DBSCAN, is a powerful tool for finding outliers by identifying various sized and shaped clusters. By using hierarchical clustering, the nested grouping of customers is seen as a dendrogram, allowing the appropriate level of hierarchy to be selected for the final segmentation procedure.

3.4 Cluster Profiling and Analysis

Each customer category is investigated to understand its unique characteristics and behaviors. In-depth profiles with demographic information, purchase patterns, and involvement levels are created for each category. Visualization techniques such as pie charts, bar charts, and radar maps are used to clearly display these segment profiles, making it easier to identify and understand the important differences and similarities across the segments. This study provides useful data that can be applied to improve marketing campaign targeting and boost customer engagement.

3.5 Implementation and Monitoring

Integrate the segmentation model with the business's customer relationship management (CRM) system or marketing platform. Make sure the segments accurately reflect the behavior of your current clients by monitoring their efficacy and applicability over time. The model needs to be updated frequently to take into account changes in customer behavior and market conditions in order to stay relevant and in line with business goals.

4. ALGORITHMS USED IN CUSTOMER SEGMENTATION

4.1 K-Means Clustering

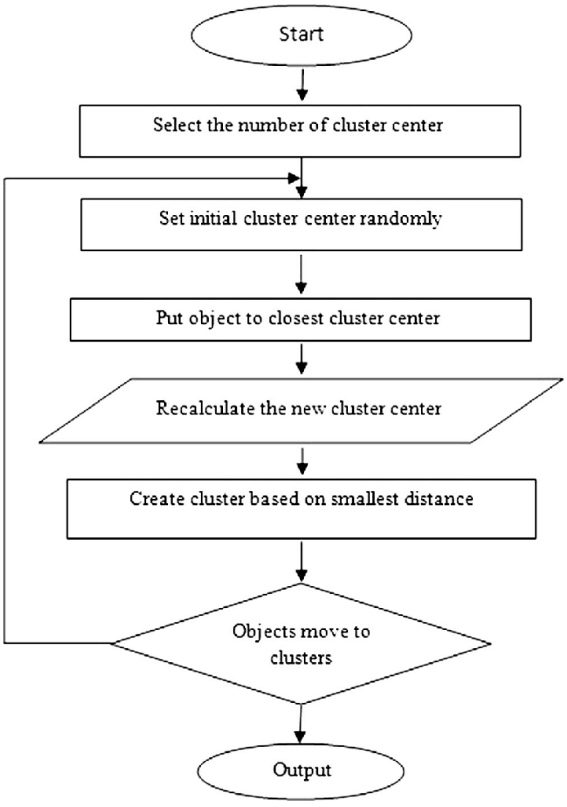


Figure: Flowchart of k-means clustering algorithm

K-Means By reducing the variation within each cluster, the clustering approach splits clients into a predetermined number of clusters. The algorithm iterates between two steps: assigning each customer to the closest centroid and updating the centroids based on the average position of customers in each cluster. The process begins by randomly picking K initial cluster centroids. Until the centroids stabilize or the allotted number of iterations is achieved, this process keeps going. The Elbow Method, which determines the point at which adding more clusters gives declining benefits, and the Silhouette Analysis, which gauges each customer's degree of similarity to their allocated cluster in relation to other clusters, can be used to calculate the ideal number of clusters (K).

4.2 Hierarchical Clustering

A tree diagram, or tree-like structure made of nested clusters, is the result of Hierarchical Clustering. There are two methods to approach it: divisive (top-down) and agglomerative (bottom-up). Agglomerative clustering involves starting each data point as its own cluster then iteratively merging pairs of clusters based on how similar they are, continuing this process until all the points are part of a single cluster or the target number of clusters is reached. Divisive clustering divides the data points into smaller clusters iteratively, beginning with all of them in a single cluster. The form and merging of clusters are dependent on the distance measure (e.g., single-linkage, complete-linkage) selected. Hierarchical clustering, which provides a comprehensive picture of the clustering process and does not entail establishing the number of clusters, may require more processing resources for large datasets.

The Algorithm Steps for Hierarchical Clustering

Step1: Construct Clusters

Agglomerative: Start with distinct clusters of individual data points.

Dividend: Assume that each data point is part of a single cluster at first.

Step2: Calculating Distances

Determine the distance matrix, which shows the distances between every two data points or clusters.

Stpe3: Merge or split Clusters

Agglomerative: Find and merge the two closest clusters. Reload the distance matrix. Continue until all data points are included in a single cluster or until the necessary number of clusters is reached.

Divisive: Find and split the cluster with the highest proportion of distinct data points. Reload the distance matrix. Continue until all data points are integrated in a single cluster or until the necessary number of clusters is reached.

Step4: Update the Distance Matrix

Recalculate and update the distances between the new clusters after splitting or merging existing clusters.

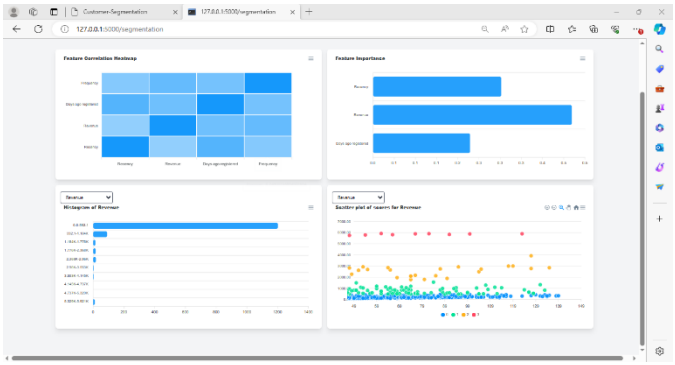
Step5: Build a Dendrogram

To see the hierarchical structure of the clusters and the order of merges and splits, make a dendrogram.

Step6: Select the Final Clusters

To cut and define the final clusters, select a level in the dendrogram.

5. RESULT



Customer segmentation based on machine learning found many groups among the clientele. Whereas DBSCAN identified clusters of different sizes and shapes and effectively handled outliers, K-Means clustering identified numerous unique segments with similar properties. Hierarchical clustering resulted in a comprehensive dendrogram that showed the hierarchical relationships between groups. Each segment was profiled based on demographic information, buying trends, and involvement levels in order to better understand and target different client groups. Visualizations such as scatter plots and heatmaps helped to further clarify the distribution and arrangement of these divides.

6. CONCLUSION

Customer segmentation has been effective in identifying distinct consumer groups with unique characteristics and behaviors by utilizing machine learning algorithms. K-Means clustering created meaningful and unambiguous segments; DBSCAN clustering discovered and addressed oddly shaped groups; and Hierarchical clustering produced a dendrogram that provided a comprehensive picture of the interconnections among the clusters. Deeper understanding of customer preferences and behavior was made possible by these segments' comprehensive profiles and graphics. This segmentation paves the way for more customer involvement, targeted marketing initiatives, and efficient company decision-making. When all is said and done, machine learning has shown to be a helpful tool for customer segmentation, assisting in more effective marketing campaign targeting and plan modification.

REFERENCES

1. T. Kansal, S. Bahuguna, V. Singh and T. Choudhury, "Customer Segmentation using K - means Clustering," International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, pp. 135-139, doi: 10.1109/CTEMS.2018.8769171.

2. Nurma Sari, Juni & Nugroho, Lukito & Ferdiana, Ridi & Santosa, Paulus. Review on Customer Segmentation Technique on Ecommerce. Advanced Science Letters. 2016.

3. T. Kanungo, d. M. Mount, n. S. Netanyahu, c. D. Piatko, r. Silverman and a. Y. Wu, "an efficient k-means clustering algorithm: analysis and implementation," in ieee transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 881-892, july 2002, doi: 10.1109/tpami.2002.1017616.

4. V. vijiles1, a. M. (2021). Customer segmentation using machine learning. International research journal of engineering and technology (irjet).

5. Chinedu Pascal, Simeon Ozuomba. "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", International Journal of Advanced Research in Artificial Intelligence, 2015.

6. Bhade, Kalyani, et al. "A Systematic Approach to Customer Segmentation and Buyer Targeting for Profit Maximization." 2018 9th International Conference onComputing, Communication and Networking Technologies (ICCCNT). IEEE, 2018.

7. Pascal, C., Ozuomba, S., & Kalu, C. (2015). Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. International Journal of Advanced Research in Artificial Intelligence, 4(10). https://doi.org/10.14569/ijarai.2015.041007.

8. Tripathi, S., A. Bhardwaj, and E. Poovammal. "Approaches to clustering in customer segmentation." International Journal of Engineering & Technology 7.3.12 (2018): 802-807.

9. Deng, Yulin, and Qianying Gao. "A study on e-commerce customer segmentation management based on improved K-means algorithm." Information Systems and e-Business Management (2018): 1-14.

10. Loh, Woong-Kee, and Young-Ho Park. "A survey on density-based clustering algorithms." Ubiquitous information technologies and applications. Springer, Berlin, Heidelberg, 2014. 775-780.