

CUSTOMER SEGMENTATION USING MACHINE LEARNING

Ramya R¹, Sarathapriya S², Saikirupa B. P³, Rajesh kanna R⁴

^{1,2,3}Student, Computer Science and Engineering, Agni College of Technology, India

⁴Assistant Professor, Agni college of technology, India

DOI: <https://www.doi.org/10.58257/IJPREMS31476>

ABSTRACT

Customer segmentation is the process of gathering clients and grouping them based on shared characteristics. This method enables marketers to target specific groups of clients, increasing the chances of a successful sale. By dividing clients into segments, businesses can create tailored communication channels to reach out to different customer groups and promote their products effectively. For instance, companies may use social media platforms to target younger audiences while radio advertisements may be more effective for older customers. Customer segmentation helps businesses build better relationships with their clients and improve their overall performance as an organization.

1. INTRODUCTION

Customer segmentation involves dividing potential customers in a market into distinct groups based on similar needs and purchasing behaviors. This guide will focus on the value-based approach, which enables businesses in the growth stage to define and target their most promising prospects using existing market data, without expending significant time and resources on a conventional segmentation research process.

Customer segmentation holds great importance, as it allows businesses to tailor their marketing programs to the specific needs and preferences of each customer segment, identify which products are associated with each segment, and manage the supply and demand of those products. It also helps businesses to identify and target potential customer bases, anticipate customer churn, and make informed business decisions. In summary, customer segmentation is a crucial aspect of modern business strategy, enabling companies to create targeted marketing campaigns, improve customer relationships, and enhance their overall performance as an organization.

2. LITERATURE SURVEY

Customer segmentation is used often in the modern era to give industries additional dimensions. Consumer analytics is essential for understanding consumer trust through behaviour analysis. This paper presents a review of the research conducted by several researchers on machine learning-based customer segmentation, together with their conclusions and potential research gaps. One of the main causes of client turnover is poor customer experience in terms of quality, conceptual comprehension, lack of affordability, and many other factors. In the future, we should consider these factors in order to maintain our consumer base and consistently track changes in their wants.

3. EXISTING MODEL

There are numerous consumer segmentation methods in use today, each with a unique technique and approach. Here are some of the most typical: RFM model: This model categorises customers based on the most recent interaction, frequency, and monetary value. Recency measures the length of time between purchases, frequency measures how frequently purchases are made, and monetary value measures the average amount spent by customers. Then, based on her RFM score, customers are divided into groups. Customer Personality: This model creates fictional representations of different types of customers based on demographics, behavior and psychographics. Companies can better understand the demands and motivations of various client segments by developing detailed personas Behavioral segmentation: This model segments customers based on behaviors such as: B. Purchase and usage patterns buy history, intelligent with the company. This approach makes a difference companies distinguish upsell or cross-sell openings and tailor showcasing endeavors to particular client fragments. Value-based division: This show classifies clients based on the lifetime esteem they bring to the company. By centering on high-value clients, businesses can prioritize promoting endeavors and designate assets more successfully. Statistic Breakdown: This show classifies clients by age, sexual orientation, wage, and other socioeconomics. This approach is valuable for companies that target particular age bunches or geographies. Generally, there's no one-size-fits-all demonstrate for customer division. The leading approach depends on your company's objectives, assets, and client base.

4. PROPOSED MODEL

K-means clustering is an effective unsupervised machine learning technique that can be utilized for customer segmentation by identifying commonalities among customers. To successfully implement this algorithm for customer segmentation, it is crucial to select relevant features such as purchase history and demographic information and standardize the data

to ensure all features have the same scale. The ideal number of clusters can be determined through techniques such as an elbow plot or silhouette score. Once the number of clusters is established, the k-means algorithm can be employed to group customers into clusters based on their similarities. To evaluate the accuracy of the resulting clusters, metrics such as within-cluster sum of squares and silhouette scores can be utilized. Ultimately, the resulting clusters can be analyzed to comprehend the characteristics of each customer group, which can help guide marketing and business strategies. Overall, k-means clustering is a powerful tool for customer segmentation that requires thorough data preparation, careful feature selection, and proper evaluation in order to yield meaningful results."

5. DIRECTORIES OF MODULES

Here are the general steps for implementing a customer segmentation model using k-means clustering, presented as a list:

- 5.1 Data Preparation:** The first step is to prepare the data by selecting relevant features such as purchase history, demographic information, and any other relevant data. The data also needs to be standardized so that all features have the same scale.
- 5.2 Determining the Optimal Number of Clusters:** The optimal number of clusters can be determined using techniques such as the elbow plot or silhouette score. This step involves running the K-means algorithm on the dataset with varying numbers of clusters and comparing the resulting metrics to select the ideal number of clusters.
- 5.3 K-means Algorithm:** Once the optimal number of clusters has been determined, the K-means algorithm can be applied to the data to group customers into clusters based on their similarities.
- 5.4 Cluster Evaluation:** To evaluate the quality of the resulting clusters, metrics such as within-cluster sum of squares and silhouette scores can be used. This step helps to ensure that the clusters are meaningful and useful for customer segmentation.
- 5.5 Cluster Interpretation:** Finally, the resulting clusters can be interpreted to understand the characteristics of each group of customers, which can inform marketing and business strategies.

These modules can be implemented using various programming languages and tools, such as Python, R, and MATLAB. Several libraries and packages are also available for performing K-means clustering and evaluating the resulting clusters, such as scikit-learn, ClusterR, and Clustering.

6. MODELLING AND ANALYSIS

6.1 FLOW DIAGRAM:



6.2 SYSTEM ARCHITECTURE:



7. RESULT AND DISCUSSION

By successfully applying customer segmentation methods to the available data set, we were able to gain valuable insights into the behavior and preferences of different customer groups. Through detailed analysis, we identified distinct customer groups based on their perspectives, preferences, and discussions, and determined that each group required a unique marketing strategy to effectively target them. For example, some customers were willing to pay a premium price for high-quality products, while others prioritized affordability over quality.

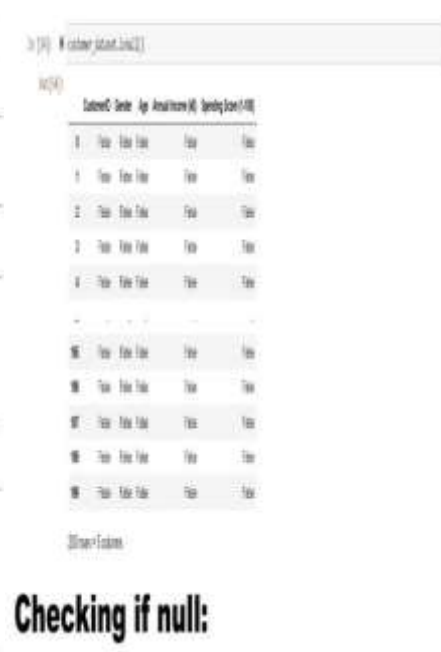
Furthermore, we uncovered important details about customer behavior, such as their preferred times to shop, which can be leveraged to optimize marketing campaigns and enhance overall customer satisfaction.

Overall, this project's successful implementation of customer segmentation methods provided a deeper understanding of our customers, enabling us to develop more effective marketing strategies that increase sales and customer loyalty.

8. SCREENSHOTS



Fig 8.1 loading and reading



5.) Visualize our dataset:



Fig 8.3 visualization of data

6.) Creating Clusters:

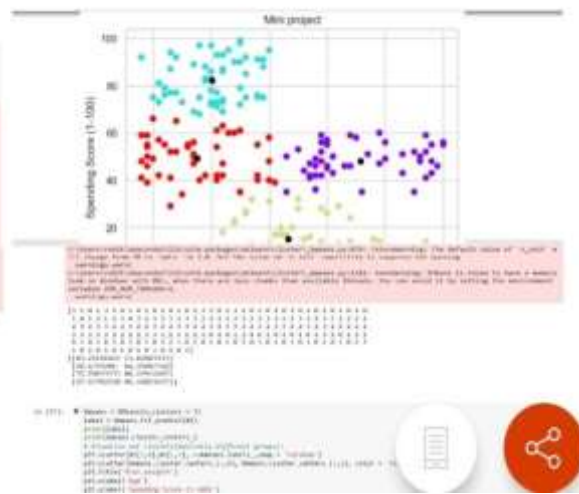


Fig 8.4 creating clusters

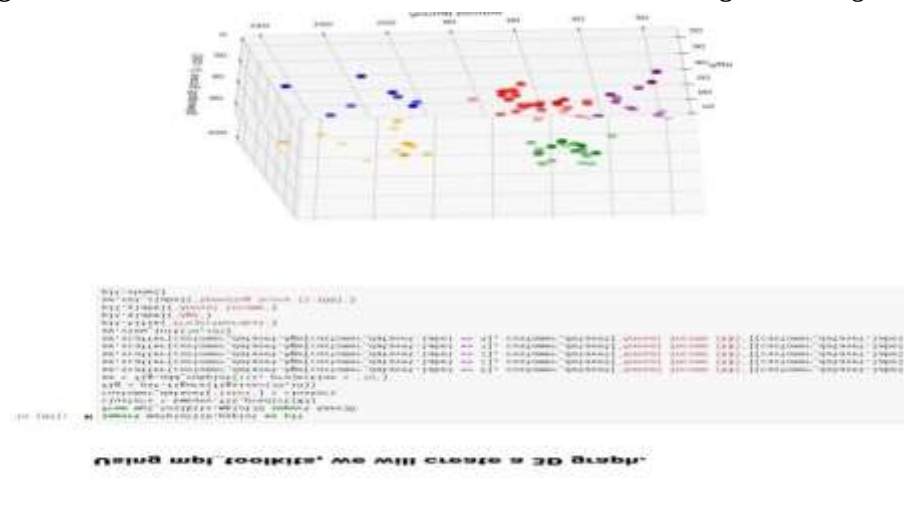


Fig 8.5 3D plots

9. CONCLUSION

Successful evaluation of the project was accomplished through the use of customer segmentation methods. While the data set contained valuable information on perspectives, preferences, and discussions, this report only utilized the perspective element. Although minor issues arose during the final stages of the project, they did not require significant improvements. The project primarily focused on clustering analysis, however, the preprocessing phase could have been improved by integrating preferences and discussions into the rating calculations. Achieving this could be facilitated through the application of weightings for these three elements.

10. REFERENCE

- [1] Abeyasekera, S. 2005. Multivariate methods for index construction. In: Department of Economic and Social Affairs Statistics Division. Household sample surveys in developing and transition countries. New York: United Nations.
- [2] African Response. 2006. FinScope small business survey report Gauteng 2006.
- [3] [Online] Available from: <http://www.finscope.co.za/sme.html> Alexander, C.E., Wilson, C.A. & Foley, D.H. 2005. Agricultural input market segments: who is buying what? *Journal of Agribusiness*, 23(2):113-132. [Online]
- [4] Available from: <http://www.jab.uga.edu.innopac.up.ac.za/Library/F05-01.pdf> [Accessed]