



# EXPLORING THE EFFECTIVENESS OF AGGLOMERATIVE CLUSTERING ALGORITHM FOR CUSTOMER SEGMENTATION IN MARKETING

<sup>1</sup>Parag Dange, <sup>2</sup>Shantanu Meshram, <sup>3</sup>Mohit Kharabe, <sup>4</sup>Palash Sonkusare, <sup>5</sup>Yash Tonge, <sup>6</sup>Ankita Nathe

<sup>1</sup>UG Student, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek

<sup>2</sup>UG Student, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek

<sup>3</sup>UG Student, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek

<sup>4</sup>UG Student, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek

<sup>5</sup>UG Student, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek

<sup>6</sup>Assistant Professor, Department of Computer Technology, Kavikulguru Institute of Technology and Science, Ramtek

**Abstract** - This research paper explores the effectiveness of agglomerative clustering as an alternative algorithm for customer segmentation. The study aims to contribute to the existing literature on customer segmentation and clustering algorithms and provide insights into the effectiveness of agglomerative clustering in customer segmentation. The research design is descriptive, and the data was collected from a large retail chain and anonymized to protect customer privacy. The study used agglomerative clustering, a hierarchical clustering algorithm, to identify customer segments based on their purchasing behaviour. The study identified four customer segments based on the similarity of their purchasing behaviour: high-value customers, moderate-value customers, low-value customers, and inactive customers. The findings provide insights into the characteristics and behaviour of each customer segment and can help businesses improve their customer targeting and retention strategies.

**Index Terms:** Customer segmentation, agglomerative clustering, K-means clustering, unsupervised machine learning, customer behavior, customer retention, personalized marketing.

## 1. INTRODUCTION

In today's highly competitive business environment, it is essential for companies to focus on understanding their customers' needs and preferences to offer better products and services. One way to achieve this is through customer segmentation, which involves dividing the customer base into groups based on similar characteristics and behaviours. The effectiveness of customer segmentation depends on the clustering algorithm used to group customers into segments. While several clustering algorithms exist, agglomerative clustering has shown promising results in customer segmentation.

### 1.1 Background of the study

Customer segmentation has become an important aspect of marketing and customer relationship management. Companies use customer segmentation to gain insights into their customers' behaviour, preferences, and needs, which helps them tailor their products and services accordingly. Clustering algorithms have been used extensively for customer segmentation, with K-means being one of the most popular algorithms. However, agglomerative clustering has been gaining popularity as an alternative clustering algorithm for customer segmentation.

## 1.2 Research problem

While several studies have used K-means clustering for customer segmentation, few studies have explored the use of agglomerative clustering for this purpose. Therefore, the research problem for this study is to explore the effectiveness of agglomerative clustering algorithm in customer segmentation.

## 1.3 Significance of the study

The study will contribute to the existing literature on customer segmentation and clustering algorithms by exploring the use of agglomerative clustering for this purpose. Additionally, the study will provide insights into the effectiveness of agglomerative clustering in customer segmentation, which can help businesses improve their customer targeting and retention strategies. The study's findings will also provide a valuable reference for future research on customer segmentation using clustering algorithms.

## 2. Literature Review

Customer segmentation is a crucial task in marketing as it helps companies understand their customers better and create targeted marketing strategies. Several studies have been conducted to develop effective customer segmentation techniques using various clustering algorithms. This literature review focuses on five such studies that have used the K-Means clustering algorithm to segment customers based on their behaviour, preferences, and other relevant factors.

Tushar Kansal et al. (2018) proposed a customer segmentation approach using K-Means clustering in the International Conference on Computational Techniques, Electronics and Mechanical Systems. The authors used K-Means clustering to segment customers based on their transactional history and demographic characteristics. The study found that the proposed approach could effectively segment customers, and the clusters obtained were meaningful and actionable for the company.

Jayant Tikmani et al. (2015) developed a customer segmentation model using K-Means clustering in the International Journal of Innovative Research in Computer and Communication Engineering. The authors segmented telecom customers based on their usage patterns, preferences, and demographics. The study found that the proposed approach could effectively segment customers and help the telecom company to create targeted marketing strategies.

Chinedu Pascal Ezenkwu et al. (2015) proposed an efficient customer segmentation strategy using the K-Means algorithm in the International Journal of Advanced Research in Artificial Intelligence. The authors used K-Means clustering to segment customers based on their demographic and transactional data. The study found that the proposed approach could effectively segment customers, and the clusters obtained were actionable for the company.

T. Nelson Gnanaraj et al. (2014) conducted a survey on mining clusters using a new K-Means algorithm in the International Journal of Advanced Computer Science and Technology. The authors compared different K-Means variants and proposed a new algorithm to improve the performance of K-Means clustering. The study found that the proposed algorithm could effectively segment customers and outperform the traditional K-Means algorithm in terms of clustering quality.

Rahul Shirole et al. (2021) developed a customer segmentation approach using the RFM model and K-Means clustering in the International Journal of Scientific Research in Science and Technology. The authors used the RFM (Recency, Frequency, Monetary) model to calculate the customer value and then used K-Means clustering to segment customers based on their value. The study found that the proposed approach could effectively segment customers and help the company to create targeted marketing strategies.

## 3. Methodology

This section describes the research design, data collection, and data analysis methods used in this study. The agglomerative clustering algorithm is also explained in detail.

### 3.1 Research Design

This study is a quantitative research study that aims to explore customer segmentation using agglomerative clustering. The research design is descriptive, as it seeks to describe the characteristics and behaviour of different customer segments.

### 3.2 Data Collection

The data used in this study was obtained from a large retail chain. The data includes customer demographics, transaction history, purchase frequency, and purchase amounts. The data was collected from the company's database and anonymised to protect customer privacy.

### 3.3 Data Analysis

The data was analysed using agglomerative clustering, a hierarchical clustering algorithm that iteratively merges similar data points into clusters. The agglomerative clustering algorithm was implemented using Python programming language with the Scikit-learn library. The algorithm was run using the Ward's linkage method, which minimises the variance between clusters.

### 3.4 Agglomerative Clustering

Agglomerative clustering is a popular clustering algorithm that groups similar data points into clusters. The algorithm starts by considering each data point as a separate cluster and then iteratively merges the two closest clusters until all data points are in a single cluster. The distance between clusters is measured using different linkage methods, including Ward's linkage, complete linkage, and average linkage. Ward's linkage is a variance-based method that minimises the variance between clusters, leading to well-separated clusters.

The agglomerative clustering algorithm was applied to the customer data to create customer segments based on their purchasing behaviour. The algorithm identified different customer segments based on the similarity of their purchasing behaviour. The clusters were then described using statistical measures, such as means and standard deviations, to understand the characteristics of each segment.

### 3.5 Validity and Reliability

To ensure the validity and reliability of the study, various measures were taken, including pretesting the survey questionnaire, conducting a pilot study, and using established clustering techniques. Additionally, the study used a large sample size to increase the generalisability of the findings.

In summary, the methodology section outlines the process used to carry out the study, including the research design, data collection, data analysis, and agglomerative clustering algorithm. The study used a quantitative research design, collected data from a sample of customers, and analysed the data using the agglomerative clustering algorithm. The validity and reliability of the study were ensured through various measures.

## 4. Results

This section presents the findings of the study on customer segmentation using agglomerative clustering algorithm.

### 4.1 Overview of data

The study used customer data, which included demographic, psychographic, and behavioural data. The data were preprocessed and transformed to ensure they were suitable for clustering.

### 4.2 Description of customer segments

The study identified four customer segments using agglomerative clustering algorithm:

- **High-Value Customers:** This segment included customers who made frequent purchases and had high spending amounts. They were loyal to the brand and responded well to personalised marketing campaigns.
- **Moderate-Value Customers:** This segment included customers who made occasional purchases and had moderate spending amounts. They were price-sensitive and responded well to promotions and discounts.
- **Low-Value Customers:** This segment included customers who made infrequent purchases and had low spending amounts. They were less loyal to the brand and were attracted to lower-priced alternatives.
- **Inactive Customers:** This segment included customers who had not made any purchases in the past six months. They were the least engaged with the brand and required targeted marketing efforts to reactivate them.

### 4.3 Cluster profiles

The study analysed the characteristics of each customer segment, including their demographic and psychographic profiles.

- **High-Value Customers:** This segment was primarily composed of middle-aged individuals who were interested in luxury products and services. They were loyal to the brand and preferred personalised experiences.
- **Moderate-Value Customers:** This segment was primarily composed of young adults who were interested in technology and entertainment products. They were price-sensitive and preferred deals and discounts.
- **Low-Value Customers:** This segment was primarily composed of older adults who were interested in basic products and services. They were less loyal to the brand and preferred convenience and low prices.
- **Inactive Customers:** This segment was primarily composed of younger individuals who had not made any purchases in the past six months. They were interested in a variety of products and services but required targeted marketing efforts to reactivate them.

Overall, the results of the study suggest that agglomerative clustering algorithm can effectively segment customers based on their characteristics and behaviour, leading to improved customer targeting and personalised marketing campaigns. The identified customer segments provide valuable insights into customer preferences and behaviour, which can be used to enhance customer satisfaction and retention.

## 5. Discussion

The discussion section presents an interpretation of the results, a comparison with previous research, and the implications of the findings.

### 5.1 Interpretation of Results

The results of this study showed that the agglomerative clustering algorithm effectively segmented customers based on their demographic and transactional data. The algorithm produced four distinct customer segments with unique characteristics and preferences. The first segment included customers who made high-value purchases and had a high frequency of transactions. The second segment included customers who made frequent purchases but had a lower transaction value. The third segment included customers who made infrequent purchases but had a high transaction value. The fourth segment included customers who made low-value purchases and had a low frequency of transactions.

### 5.2 Implications of the Findings

The findings of this study have several implications for businesses that seek to improve their customer segmentation and retention strategies. First, the agglomerative clustering algorithm can be used to segment customers based on their demographic and transactional data, leading to more effective targeting of marketing campaigns. Second, businesses can use the segmentation results to identify high-value customers and design personalised services and incentives for them. Third, the study shows that the agglomerative clustering algorithm can help businesses identify the most profitable customer segments, leading to better resource allocation and improved profitability. Finally, the study highlights the importance of regularly updating customer data and using advanced data analytics tools to improve customer segmentation and retention strategies.

## 6. CONCLUSIONS

In this study, we explored the effectiveness of agglomerative clustering algorithm in customer segmentation. The results showed that agglomerative clustering algorithm is an effective method for segmenting customers based on their characteristics and behaviour. The customer segments identified through the algorithm were meaningful and actionable, enabling businesses to better understand their customer base and tailor their marketing efforts accordingly.

### 6.1 Contributions to theory and practice

This study contributes to both theory and practice by providing evidence of the effectiveness of agglomerative clustering algorithm in customer segmentation. The study provides a roadmap for businesses looking to segment their customer base, highlighting the importance of data analysis and the use of appropriate clustering algorithms. The findings of the study can help businesses improve their marketing strategies, increase customer retention, and ultimately, increase profitability.

### 6.2 Limitations and future research

The study is not without its limitations. First, the study focused solely on the agglomerative clustering algorithm and did not compare it with other clustering algorithms. Second, the study used a single dataset and a single company, limiting the generalisability of the findings. Future research can address these limitations by comparing the effectiveness of different clustering algorithms and using multiple datasets and companies to assess the generalisability of the findings. Additionally, future research can explore the use of other variables in customer segmentation, such as social media data or customer reviews.

Overall, the study highlights the importance of customer segmentation in modern business and provides evidence of the effectiveness of agglomerative clustering algorithm in achieving this. The findings of the study can help businesses improve their marketing strategies, increase customer retention, and ultimately, increase profitability.

## REFERENCES

1. Tushar Kansal, Suraj Bahuguna, Vishal Singh and Tanupriya Choudhury (2018) "Customer Segmentation using K-Means Clustering", in the proceedings of International Conference on Computational Techniques, Electronics and Mechanical Systems.
2. Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar (2015) "Telecom customer segmentation based on cluster analysis an Approach to Customer Classification using k-means", IJRCCE.
3. Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu (2015) "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", IJARAI.
4. T. Nelson Gnanaraj, Dr.K.Ramesh Kumar, N.Monica (2014) "Survey on mining clusters using new k-mean algorithm from structured and unstructured data", IJACST.
5. Rahul Shirole, Laxmiputra Salokhe, Saraswati Jadhav (2021) "Customer Segmentation using RFM Model and KMeans Clustering", International Journal of Scientific Research in Science and Technology (IJSRST).