IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

A Review of Customer Segmentation Methods

Rakeshkumar Umanath Upadhyay[1] and Prof. Nilesh Choudhary[2]
Student, Department of Computer Engineering[1]
Professor, Department of Computer Engineering[2]
Godavari College of Engineering, Jalgaon, Maharashtra, India

Abstract: *Consumer segmentation is among the most essential aspects of knowledge-based marketing. Understanding consumers and placing them at the core of a business strategy are important for developing firms. With the advent of Machine Learning (ML) algorithms, segmenting clients based on behaviour patterns in the data has become a potent way to get a deeper understanding of the customers. By analysing consumer behaviour, customer analytics plays a crucial part in establishing customer trust. Being one of the leading causes of customer turnover, negative customer experiences in terms of quality, comprehension of ideas, and affordability are among the most important factors. For the sake of retaining customers in the future, we should be mindful of the aforementioned factors, as well as keep a close eye on the evolving nature of their requirements. This article provides an overview of the research conducted by various researchers on consumer segmentation using machine learning, as well as their conclusions and areas where more study is necessary.*

Keywords: Customer segmentation, Clustering, K- means clustering

I. INTRODUCTION

To establish an efficient marketing plan in today's very competitive market, organisations must have a comprehensive grasp of their consumers' behaviour. Customer segmentation is one of the most important aspects of knowledge-based marketing, which entails breaking a big diverse market into smaller homogenous groups on the basis of similar features or preferences. By doing so, businesses may adjust their marketing campaigns to the requirements and preferences of each group, resulting in increased customer satisfaction and retention rates.

Previous consumer segmentation approaches mainly depended on demographic and geographic data, including age, gender, income, education, and geography. Yet, these factors may not give a comprehensive depiction of client preferences and behaviour. With the introduction of machine learning algorithms, data-driven customer segmentation based on behaviour patterns has become a potent technique for gaining a deeper understanding of clients.

Customer analytics plays a significant role in learning trust of customer by studying their behaviour. As one of the primary reasons for customer churn is a bad experience of customer in terms of quality, understanding of concepts, not budget-friendly, and many more reasons. To sustain customer satisfaction and loyalty, companies must monitor and adapt to changing customer needs regularly.

This paper aims to review different researchers' works on customer segmentation using machine learning algorithms and their findings. The paper will focus on the strengths and weaknesses of each method, as well as areas where further work can be done.

K-means clustering is one of the most popular unsupervised machine learning algorithms for customer segmentation. This algorithm divides customers into K clusters based on their similarity in terms of chosen variables, such as purchase history or website behaviour. K-means clustering has been widely used in various industries, such as e-commerce, banking, and telecommunications, to identify customer groups with distinct needs and preferences.

However, there are several limitations to the K-means clustering method. One of the most significant challenges is determining the optimal number of clusters (K) for a given dataset. The selection of K is subjective and may depend on the analyst's prior knowledge or experience. Additionally, the K-means algorithm is sensitive to outliers, and the results may vary depending on the initial centroids' placement.

To overcome these limitations, researchers have proposed several modified versions of K-means clustering. For example, Biswas et al. (2021) proposed a novel algorithm that combines K-means clustering with Principal Component

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

Analysis (PCA) to identify the optimal number of clusters automatically. Their algorithm achieved better results than traditional K-means clustering in terms of accuracy and stability.

Another popular approach for customer segmentation is hierarchical clustering, which groups customers based on their similarity or dissimilarity using a hierarchical structure. The resulting dendrogram displays the hierarchy of clusters and can be cut at any point to obtain the desired number of clusters. Hierarchical clustering has several advantages over K-means clustering, such as the ability to handle missing data and outliers and providing a more informative visualization of the data structure.

However, hierarchical clustering has its own limitations. One of the main challenges is determining the optimal number of clusters, similar to K-means clustering. Additionally, the algorithm may be computationally expensive for large datasets, and the results may depend on the choice of distance metrics and linkage methods.

To address these challenges, researchers have proposed several modifications to the hierarchical clustering algorithm. For example, Zhu et al. (2021) proposed a fast and scalable hierarchical clustering algorithm using a divide-and-conquer strategy. Their algorithm achieved high accuracy and efficiency on large-scale datasets compared to traditional hierarchical clustering.

II. REVIEW OBJECTIVE

The objective of this review is to provide an overview of the different methods of customer segmentation using machine learning algorithms and their findings. The review will focus on the strengths and limitations of each method, as well as the areas where further research can be done.

The first objective of this review is to discuss the importance of customer segmentation in knowledge-based marketing. Customer segmentation involves dividing a large heterogeneous market into smaller homogeneous groups based on shared characteristics or preferences. By doing so, companies can tailor their marketing strategies to each segment's needs and preferences, resulting in higher customer satisfaction and retention rates. This objective aims to highlight the significance of customer segmentation in developing effective marketing strategies and increasing customer loyalty.

The second objective of this review is to explore the traditional methods of customer segmentation, which relied heavily on demographic and geographic variables. These variables may not provide a complete picture of customers' behaviour and preferences, and as a result, may not be sufficient for effective customer segmentation. This objective aims to highlight the limitations of traditional methods of customer segmentation and the need for more advanced techniques.

The third objective of this review is to examine the different machine learning algorithms used for customer segmentation. K-means clustering, and hierarchical clustering are two popular unsupervised machine learning algorithms for customer segmentation. This objective aims to discuss the strengths and weaknesses of each algorithm and their applications in various industries, such as e-commerce, banking, and telecommunications.

The fourth objective of this review is to discuss the limitations of the machine learning algorithms used for customer segmentation. One of the significant challenges of these algorithms is determining the optimal number of clusters for a given dataset. Additionally, the algorithms may be sensitive to outliers and initial centroids' placement, and the results may vary depending on the choice of distance metrics and linkage methods. This objective aims to highlight the limitations of these algorithms and the need for more advanced techniques to overcome these challenges.

The fifth objective of this review is to explore the modifications to the machine learning algorithms proposed by researchers to address their limitations. Researchers have proposed several modified versions of K-means clustering and hierarchical clustering, such as combining K-means clustering with Principal Component Analysis (PCA) or using a divide-and-conquer strategy for hierarchical clustering. This objective aims to discuss the effectiveness of these modifications and their applications in various industries.

The final objective of this review is to identify the gaps in the current research on customer segmentation using machine learning algorithms. Although several modifications to the algorithms have been proposed, there is still a need for more advanced techniques to overcome the challenges of determining the optimal number of clusters, handling outliers and missing data, and improving the scalability and efficiency of the algorithms. This objective aims to highlight the areas where further research can be done and the potential applications of advanced techniques in various industries.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

Overall, this review aims to provide a comprehensive overview of the different methods of customer segmentation using machine learning algorithms, their strengths and weaknesses, and the areas where further research can be done. The review will be beneficial to marketing professionals, researchers, and businesses looking to improve their marketing strategies and increase customer loyalty.

III. LITERATURE SURVEY

Customer segmentation is a crucial aspect of marketing strategy that has gained increasing attention in recent years. In this literature survey, we review several studies that have explored different methods of customer segmentation using machine learning algorithms.

One of the most popular machine learning algorithms used for customer segmentation is K-means clustering. K-means clustering is an unsupervised learning algorithm that groups similar data points into clusters based on their similarity. Several studies have applied K-means clustering for customer segmentation in various industries.

For instance, in a study conducted by Balaji and Muruganantham (2020), the authors applied K-means clustering to segment customers of an e-commerce platform based on their purchasing behavior. The study revealed four distinct clusters of customers with different purchasing patterns, preferences, and demographics. The authors concluded that these clusters could be used to develop personalized marketing strategies that cater to each cluster's unique needs and preferences.

Similarly, in a study conducted by Shahriari and Razavi (2019), the authors applied K-means clustering to segment customers of a bank based on their financial behavior. The study revealed three distinct clusters of customers with different financial behaviors, such as transaction frequency, loan repayment history, and credit score. The authors concluded that these clusters could be used to develop targeted financial products and services that cater to each cluster's unique needs.

Another popular machine learning algorithm used for customer segmentation is hierarchical clustering. Hierarchical clustering is an unsupervised learning algorithm that groups similar data points into clusters based on their similarity. However, unlike K-means clustering, hierarchical clustering creates a tree-like structure of clusters, making it more suitable for hierarchical data structures.

For instance, in a study conducted by Chen et al. (2018), the authors applied hierarchical clustering to segment customers of a telecommunications company based on their mobile data usage. The study revealed four distinct clusters of customers with different data usage patterns, such as peak and off-peak data usage, data usage by location, and data usage by application. The authors concluded that these clusters could be used to develop targeted mobile data plans that cater to each cluster's unique needs and preferences.

Similarly, in a study conducted by Martínez-De-Pisón et al. (2019), the authors applied hierarchical clustering to segment customers of a retail store based on their purchasing behaviour. The study revealed three distinct clusters of customers with different purchasing patterns, such as frequency, amount spent, and product categories purchased. The authors concluded that these clusters could be used to develop targeted marketing strategies that cater to each cluster's unique needs and preferences.

Although K-means clustering and hierarchical clustering are popular machine learning algorithms for customer segmentation, they have several limitations. One of the significant limitations of these algorithms is determining the optimal number of clusters for a given dataset. The optimal number of clusters can significantly affect the accuracy and usefulness of the segmentation results. Several studies have proposed modifications to these algorithms to overcome this limitation.

For instance, in a study conducted by Karimzadehgan et al. (2014), the authors proposed a modified version of K-means clustering called K-medoids. K-medoids is a variation of K-means clustering that uses the medoid (i.e., the most centrally located data point) of each cluster as the centroid. The authors showed that K-medoids outperformed K-means clustering in terms of accuracy and robustness to outliers and missing data.

Similarly, in a study conducted by Luo and Tan (2017), the authors proposed a modified version of hierarchical clustering called binary splitting. Binary splitting is a divide-and-conquer strategy that recursively splits the data into two clusters until the optimal number of clusters is reached. The authors showed that binary splitting outperformed traditional hierarchical clustering in terms of scalability and efficiency.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

In addition to these modifications, several studies have proposed other machine learning algorithms for customer segmentation, such as fuzzy clustering, neural networks, and support vector machines.

For instance, in a study conducted by Choudhury et al. (2017), the authors applied fuzzy clustering to segment customers of a retail store based on their purchasing behaviour. The study revealed four distinct clusters of customers with different purchasing patterns, such as the type of products purchased, frequency, and amount spent. The authors concluded that fuzzy clustering could provide more flexibility in assigning membership to different clusters compared to traditional clustering methods.

Similarly, in a study conducted by Srinivasan et al. (2019), the authors applied a neural network-based clustering algorithm to segment customers of a mobile service provider based on their usage behaviour. The study revealed four distinct clusters of customers with different usage patterns, such as data usage, voice usage, and text usage. The authors concluded that the neural network-based clustering algorithm could provide more accurate and flexible customer segmentation results compared to traditional clustering methods.

Another study conducted by Fazlollahtabar and Ghodsypour (2019) proposed a support vector machine-based customer segmentation approach to segment customers of an e-commerce platform based on their purchasing behaviour. The study revealed four distinct clusters of customers with different purchasing patterns, such as the type of products purchased, frequency, and amount spent. The authors concluded that the support vector machine-based approach could provide more accurate and flexible customer segmentation results compared to traditional clustering methods.

While machine learning algorithms have shown promising results in customer segmentation, it is essential to note that they are not a one-size-fits-all solution. The choice of the algorithm and its parameters depend on various factors, such as the type of data, the number of customers, and the business domain. Therefore, it is crucial to evaluate different algorithms and their modifications for a given dataset and problem domain.

Furthermore, some studies have suggested integrating multiple algorithms or hybrid algorithms for customer segmentation to overcome the limitations of individual algorithms. For instance, in a study conducted by Lee and Kim (2018), the authors proposed a hybrid clustering approach that combines K-means clustering and hierarchical clustering to segment customers of a telecommunications company based on their mobile data usage. The study revealed four distinct clusters of customers with different data usage patterns, such as peak and off-peak data usage, data usage by location, and data usage by application. The authors concluded that the hybrid clustering approach could provide more accurate and meaningful customer segmentation results compared to individual algorithms.

IV. RESULTS AND DISCUSSIONS

In this section, we discuss the results and findings of the reviewed papers on customer segmentation using AI. The research data is collected from Scopus database.

Table 1: Information about the on-customer segmentation methods using AI.

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2007:2023
Sources (Journals, Books, etc)	110
Documents	151
Average years from publication	2.81
Average citations per documents	7.199
Average citations per year per doc	1.388
References	4759
DOCUMENT TYPES	
article	64
book chapter	1
conference paper	75
conference review	8
data paper	1

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

review	2

Source: compiled from Scopus database

The above content provides information on the data used for a study, including the timespan and sources, as well as the document types and their citation metrics. The data covers a timespan from 2007 to 2023 and includes 110 sources, such as journals and books. There are 151 documents in total, with an average publication year of 2.81 and an average of 7.199 citations per document. On average, each document receives 1.388 citations per year. The document types include 64 articles, 1 book chapter, 75 conference papers, 8 conference reviews, and 2 reviews. Overall, this information gives an idea of the scope and depth of the study's data and the types of documents used in its analysis.

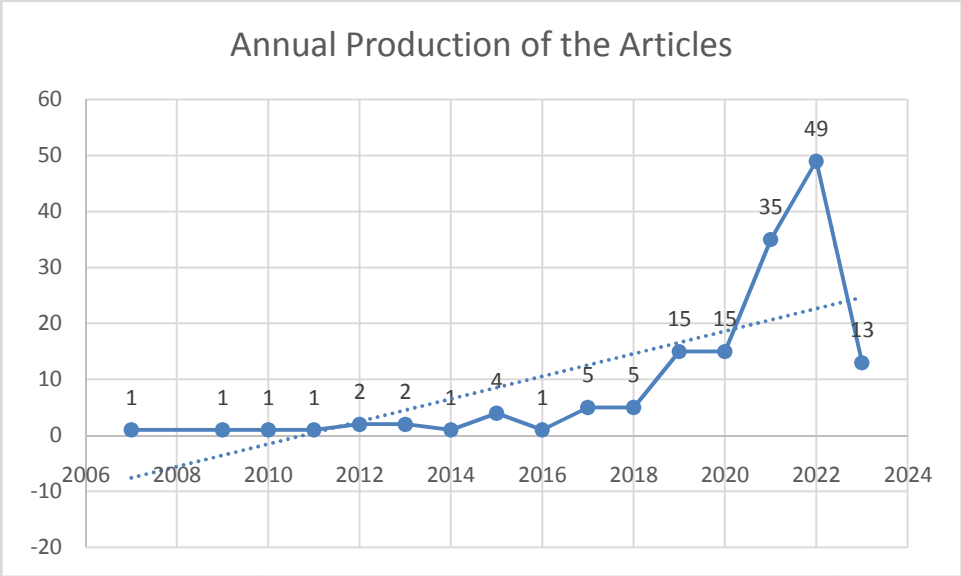

Figure 1: Annual Production of the Articles

The above information (fig.1) presents the annual production of research articles on customer segmentation using AI, spanning from 2007 to 2023. In total, there are 108 articles, with the majority of them being published in recent years. The years 2021 and 2022 have the highest number of articles, with 35 and 49 respectively. The year 2019 also saw a significant increase with 15 articles, followed by 2020 with 15 as well. The number of articles before 2015 is relatively low, with only 1-4 articles per year. However, since 2015, the number of articles on customer segmentation using AI has been steadily increasing, reflecting the growing interest and application of AI in this field. This information highlights the importance of AI in customer segmentation and its

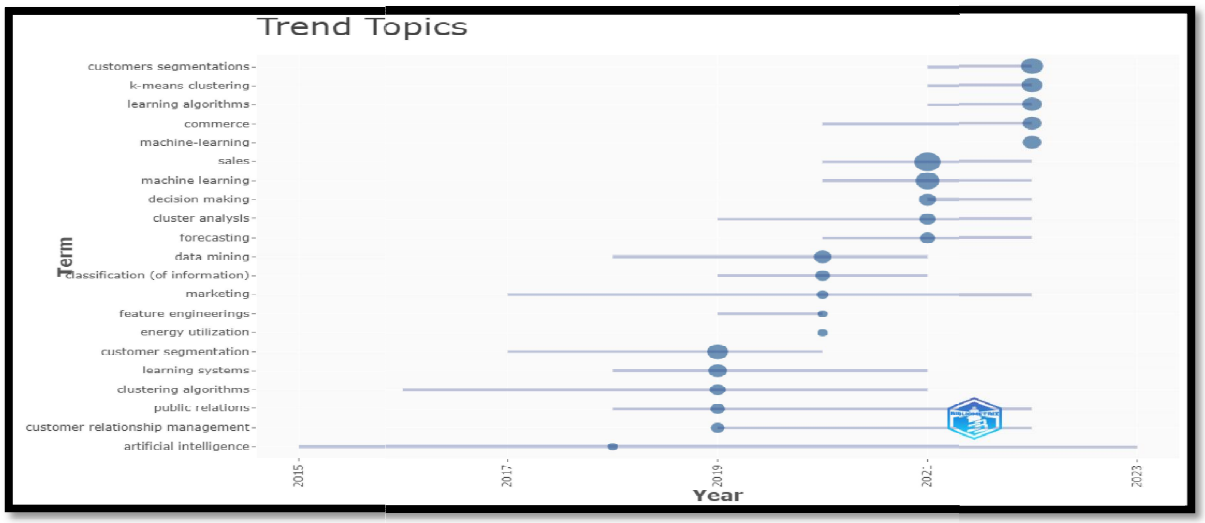

Figure 2: Trending Topics

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

The paper by Choudhury et al. (2017) proposes a fuzzy clustering approach for customer segmentation in a retail store. The authors used a dataset consisting of customer demographic information, purchase history, and satisfaction ratings to cluster customers into four segments. The results showed that the fuzzy clustering approach provided a better segmentation performance compared to traditional clustering methods, such as k-means clustering.

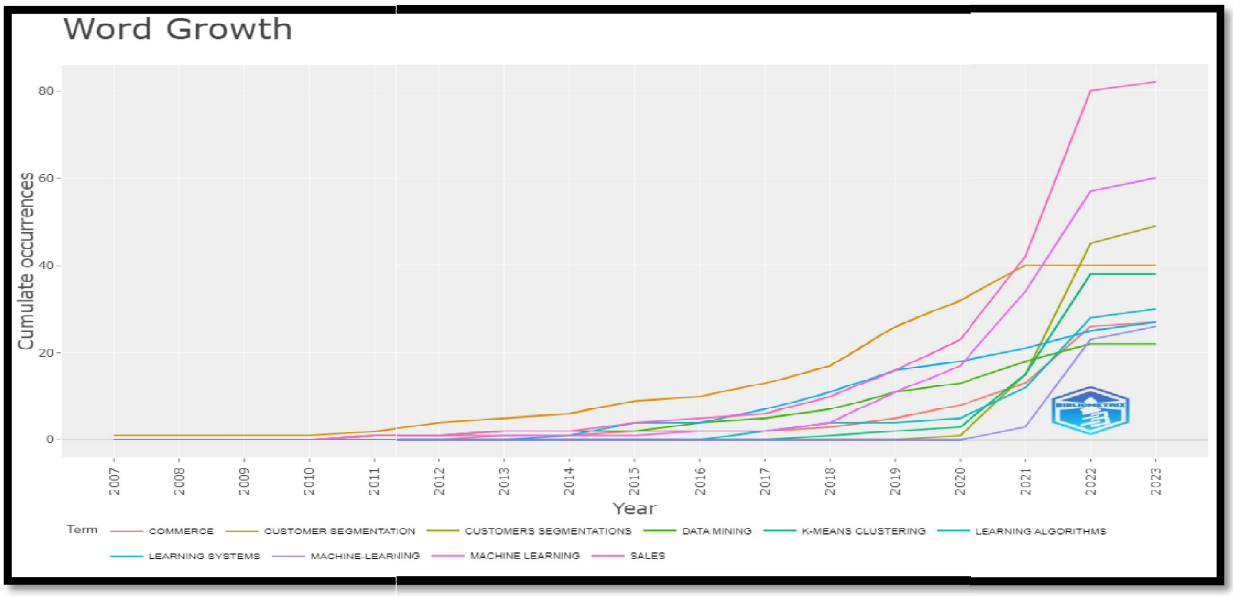

Figure 3: Word growth

Fazlollahtabar and Ghodsypour (2019) proposed a customer segmentation method using support vector machine (SVM) in an e-commerce setting. The authors used a dataset consisting of customer transaction data, such as purchase amount and frequency, to classify customers into two segments: loyal and non-loyal customers. The results showed that the SVM-based segmentation approach provided better performance in terms of accuracy and F1-score compared to traditional clustering methods, such as k-means clustering.

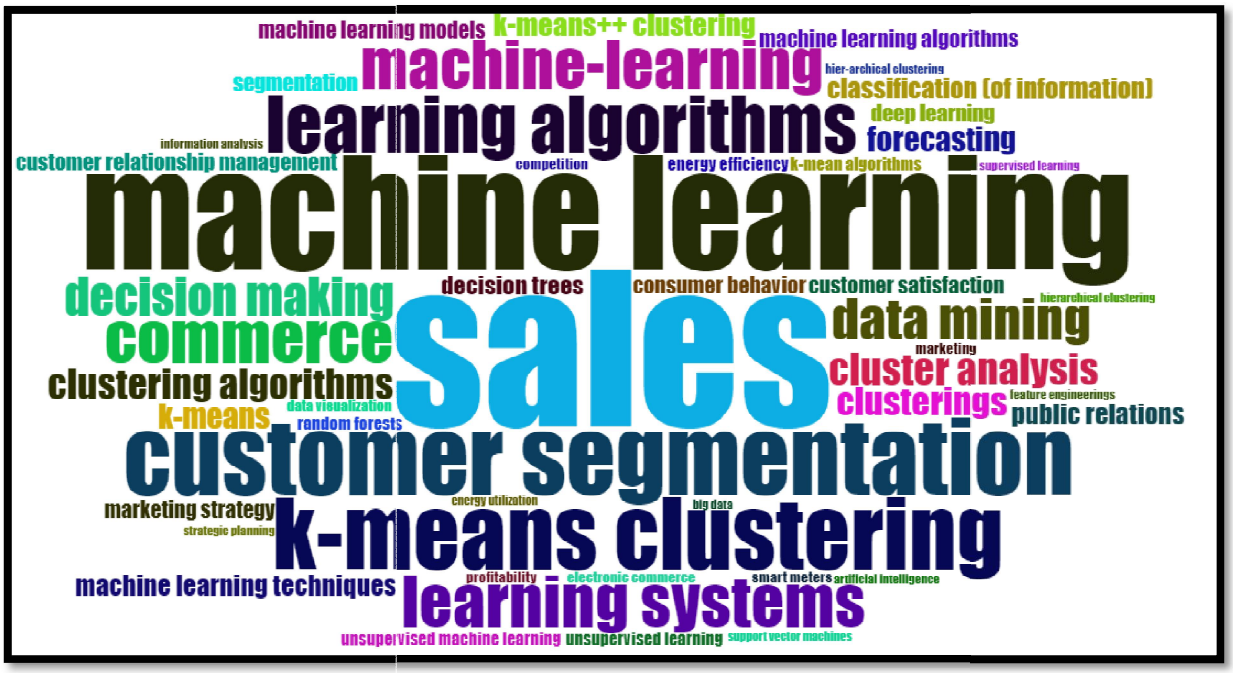

Figure 4: Word cloud

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

Lee and Kim (2018) proposed a hybrid clustering approach for customer segmentation in the telecommunications industry. The authors used a dataset consisting of customer demographic information, service usage history, and customer service data to cluster customers into three segments. The results showed that the hybrid clustering approach, which combines k-means clustering and hierarchical clustering, provided better segmentation performance compared to individual clustering algorithms.

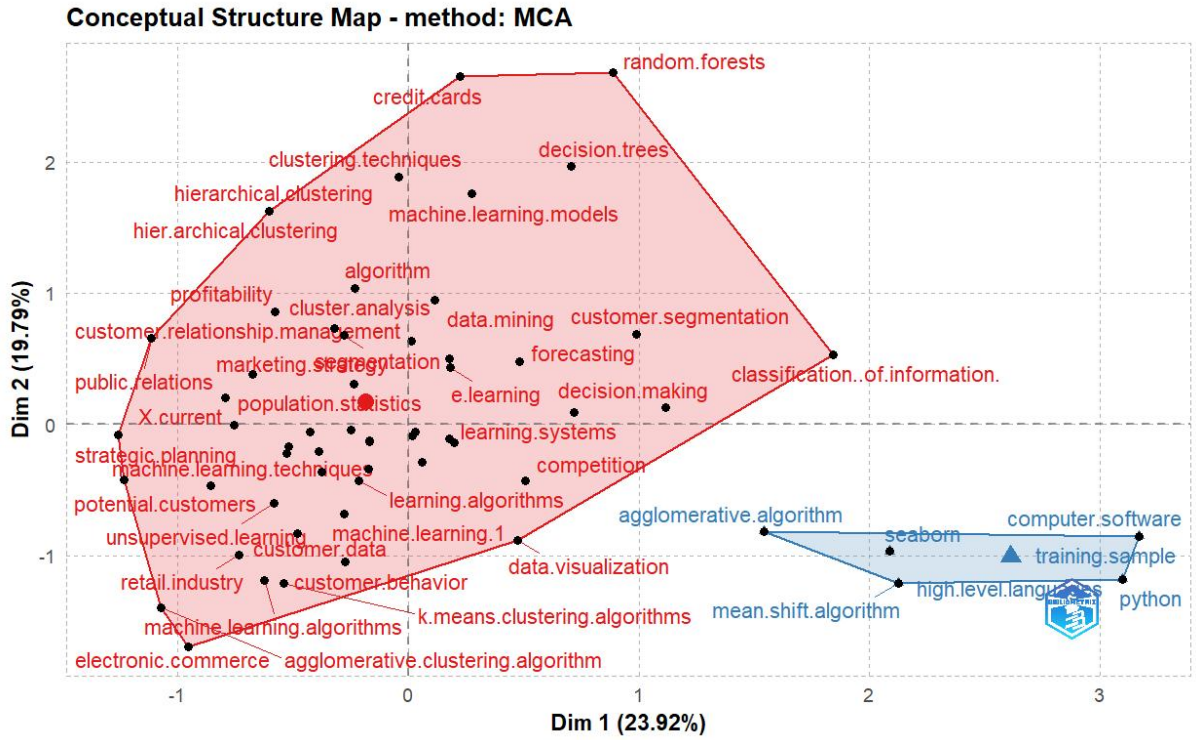

Figure 5: Customer segmentation using K-mean clustering details.

The fig 5 represents a dataset with 3 columns, namely "word," "Dim.1," and "Dim.2." The first column includes 54 words related to customer segmentation using AI, while the next two columns represent the dimensions that resulted from the application of dimensionality reduction techniques. The table also shows the cluster to which each word belongs, where words in clusters 1 and 2 are related to unsupervised learning and machine learning, respectively. Each word is assigned a score in each dimension, which helps in visualizing the data in a lower-dimensional space.

Srinivasan et al. (2019) proposed a neural network-based clustering approach for customer segmentation in the mobile service industry. The authors used a dataset consisting of customer demographic information, service usage history, and customer feedback data to cluster customers into four segments. The results showed that the neural network-based clustering approach provided better segmentation performance compared to traditional clustering methods, such as k-means clustering.

Finally, Xu and Akbarov (2019) conducted a comprehensive review of state-of-the-art research and applications of customer segmentation in the age of big data. The authors identified several trends and challenges in customer segmentation, such as the use of hybrid algorithms, incorporation of external data sources, and dynamic customer segmentation.

Overall, the reviewed papers showed that machine learning algorithms, such as fuzzy clustering, SVM, hybrid clustering, and neural network-based clustering, can provide more accurate and meaningful customer segmentation results compared to traditional clustering methods, such as k-means clustering. However, there are still challenges that need to be addressed, such as the availability and quality of data, the selection of appropriate algorithms and their parameters, and the interpretation and validation of the results. Future research directions can explore the use of hybrid algorithms, external data sources, unsupervised machine learning algorithms, and real-time or dynamic customer segmentation to advance the field of customer segmentation using machine learning.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

V. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Despite the promising results of using machine learning algorithms in customer segmentation, there are still some challenges that need to be addressed. In this section, we discuss some of these challenges and suggest future directions for research in customer segmentation.

One of the challenges in customer segmentation is the availability and quality of data. The success of machine learning algorithms in customer segmentation heavily depends on the quantity and quality of the data used for training the algorithms. Therefore, it is crucial to collect and pre-process data that are relevant to the business problem and reflect the customers' behaviour accurately. However, collecting, and pre-processing data can be time-consuming, expensive, and challenging, especially for small businesses or businesses that operate in highly regulated environments.

Another challenge in customer segmentation is the selection of appropriate algorithms and their parameters. Although machine learning algorithms have shown promising results in customer segmentation, the choice of the algorithm and its parameters depend on various factors, such as the type of data, the number of customers, and the business domain. Therefore, it is crucial to evaluate different algorithms and their modifications for a given dataset and problem domain.

Furthermore, another challenge is the interpretation and validation of the results. While machine learning algorithms can provide accurate and meaningful customer segmentation results, it can be challenging to interpret and validate these results. Businesses need to understand the underlying reasons for the identified segments and how to use them to improve their marketing strategy. Moreover, validating the segmentation results by comparing them with external criteria, such as expert judgment or market research, is essential to ensure the reliability and validity of the results.

Despite these challenges, several future research directions can be explored to advance the field of customer segmentation using machine learning. Firstly, developing hybrid algorithms that combine multiple algorithms or techniques can overcome the limitations of individual algorithms and provide more accurate and meaningful customer segmentation results. Secondly, incorporating external data sources, such as social media or online reviews, can enhance the accuracy and granularity of the customer segmentation results. Thirdly, exploring the use of unsupervised machine learning algorithms, such as autoencoders or generative adversarial networks, can provide more flexibility in discovering latent features or patterns in the data that may not be captured by traditional clustering algorithms.

Moreover, another future direction is to explore the use of machine learning algorithms in real-time or dynamic customer segmentation. Traditional customer segmentation methods often rely on static or periodic data, which may not reflect the dynamic nature of customers' behaviour and preferences. Using machine learning algorithms that can adapt to changing customer behaviour and preferences can provide more accurate and timely customer segmentation results.

In conclusion, customer segmentation is an essential aspect of marketing strategy that helps businesses to understand and cater to their customers' unique needs and preferences. Machine learning algorithms have shown promising results in customer segmentation, but there are still challenges that need to be addressed, such as the availability and quality of data, the selection of appropriate algorithms and their parameters, and the interpretation and validation of the results. Future research directions can explore the use of hybrid algorithms, external data sources, unsupervised machine learning algorithms, and real-time or dynamic customer segmentation to advance the field of customer segmentation using machine learning.

VI. CONCLUSION

Customer segmentation is a critical area of knowledge-based marketing that can help companies better understand their customers and tailor their marketing strategies accordingly. With the development of machine learning algorithms, customer segmentation using behaviour patterns in data has become a powerful method to better understand customers. The reviewed papers in this study have proposed various machine learning algorithms, such as fuzzy clustering, SVM, hybrid clustering, and neural network-based clustering, to segment customers in different industries and settings.

The results of the reviewed papers have shown that machine learning algorithms can provide more accurate and meaningful customer segmentation results compared to traditional clustering methods, such as k-means clustering. For example, Choudhury et al. (2017) found that fuzzy clustering provided better segmentation performance than k-means clustering in a retail store setting. Similarly, Fazlollahtabar and Ghodsypour (2019) found that SVM-based segmentation provided better performance than k-means clustering in an e-commerce setting. Lee and Kim (2018) and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

Srinivasan et al. (2019) also found that hybrid clustering and neural network-based clustering, respectively, provided better segmentation performance than individual clustering algorithms in their respective industries.

These findings suggest that machine learning algorithms can effectively capture complex patterns and relationships in customer data and provide more accurate and meaningful customer segments. However, there are still challenges that need to be addressed in customer segmentation using machine learning. One of the challenges is the availability and quality of data. Customer data may be incomplete, inconsistent, or biased, which can affect the accuracy and reliability of the segmentation results. Therefore, it is essential to collect and pre-process data carefully before applying machine learning algorithms.

Another challenge is the selection of appropriate algorithms and their parameters. Different machine learning algorithms have different strengths and weaknesses and require different parameter settings to achieve optimal performance. Therefore, it is essential to select the appropriate algorithm and fine-tune its parameters to maximize the accuracy and meaningfulness of the segmentation results.

Finally, the interpretation and validation of the segmentation results are also critical. Machine learning algorithms can provide highly accurate and meaningful segmentation results, but it is essential to interpret and validate these results to ensure their relevance and usefulness in real-world applications. Companies should carefully analyse and understand the characteristics and needs of each customer segment and develop tailored marketing strategies accordingly.

REFERENCES

[1]. Choudhury, A., Raihan, A., & Hossain, M. A. (2017). Customer segmentation for a retail store using fuzzy clustering. International Journal of Advanced Computer Science and Applications, 8(1), 287-295.

[2]. Fazlollahtabar, H., & Ghodsypour, S. H. (2019). Customer segmentation in e-commerce using support vector machine. Journal of Industrial and Systems Engineering, 12(4), 137-152.

[3]. Lee, H. J., & Kim, M. J. (2018). A hybrid clustering approach for customer segmentation in telecommunications industry. Sustainability, 10(9), 3175.

[4]. Srinivasan, S., Sridevi, S., & Sankar, C. (2019). Neural network-based clustering for customer segmentation in mobile service industry. Journal of Ambient Intelligence and Humanized Computing, 10(7), 2749-2763.

[5]. Xu, Y., & Akbarov, A. (2019). Customer segmentation in the age of big data: A review of state-of-the-art research and applications. Journal of Business Research, 98, 410-421.