Jurnal ELTIKOM:
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 8, No. 1, June 2024, page. 83-90
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v8i1.1137
Available online at http://eltikom.poliban.ac.id

K-MEANS CLUSTERING METHOD FOR CUSTOMER SEGMENTATION BASED ON POTENTIAL PURCHASES

Baiq Nikum Yulisasih[1,*], Herman[1], Sunardi[2]
[1)] Departement of of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2)] Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
e-mail: 2307048005@webmail.uad.ac.id, hermankaha@mti.uad.ac.id, sunardi@mti.uad.ac.id

Received: 15 April 2024 – Revised: 11 June 2024 – Accepted: 15 June 2024

ABSTRACT

The rapid growth in customer data has driven companies to develop smarter and more effective marketing strategies. One efficient approach is customer segmentation, which involves dividing a market or group of customers into smaller segments based on similar characteristics or behaviors. Customer segmentation improves understanding of customer needs, preferences, and behavior. This study uses customer segmentation based on purchase potential at Fast Moving Consumer Goods (FMCG). Analyzing potential purchases can help identify market opportunities, implement more effective pricing, target promotions, manage stock and distribution, and develop new products to enhance customer satisfaction. The most commonly used segmentation method is the K-Means Clustering algorithm, which groups data into homogeneous clusters. This study aims to segment customers based on potential purchases using the K-Means Clustering method. The customer dataset in FMCG stores was divided into three clusters using seven attributes: Sex, Marital Status, Age, Education, Income, Occupation, and Settlement Size. The results, calculated in Microsoft Excel, concluded after four iterations with three clusters: k1 (Cluster 1) with 535 customers having low purchase potential, k2 (Cluster 2) with 685 customers having high purchase potential, and k3 (Cluster 3) with 7810 customers having medium purchase potential.

Keywords*: customer segmentation, k-means clustering, potential purchases.*

I. INTRODUCTION

In an era of increasingly complex and competitive business, understanding customers profoundly is crucial for companies to achieve a competitive advantage. One significant approach is customer segmentation, which allows companies to divide the market into separate groups with similar characteristics, enabling more focused and effective marketing strategies [1]. The main objective of customer segmentation is to design and build a clustering model for customer segmentation using the k-means clustering algorithm. The urgency of this research lies in finding alternative solutions to group customer segmentation based on purchasing potential, optimizing customer retention and attracting new customers. The novelty of this study is the use of the k-means clustering algorithm to determine customer segmentation based on purchasing potential through clustering for customer profile grouping [2].

Some consumers who previously shopped at supermarkets have switched to buying products online through e-commerce platforms [3], [4]. This change in shopping patterns is driven by several reasons, such as the convenience of online shopping that can be done anytime and anywhere, a wider selection of products, and attractive promotions offered by e-commerce platforms [5], [6]. Ipsos Global Trends 2021 Aftershocks and Continuity research shows that online shopping transactions in Indonesia rank highest among 25 countries surveyed, with 73% of total transactions occurring online [7]. E-commerce transactions can be categorized into business-to-business (B2B) for large transactions between companies and business-to-consumer (B2C) for smaller transactions dominated by retail sales [8], [9].

Currently, one popular form of B2C is the sale of goods through marketplaces such as Alibaba, Shopee, and Tokopedia [10]. These B2C e-commerce companies compete fiercely to capture customers and increase their sales [11]. Meanwhile, online customers are usually very fluid and changeable both

in deciding the details of the goods they will buy and where they buy them. This is because there are so many choices in the online marketplace [12]. Extensive research, both academic and industrial, has been conducted to examine online consumer behavior [13]. Previous studies concluded that several interrelated factors influence customer decisions to buy items on an e-commerce platform [14].

These factors are known as customer preference factors. Unfortunately, there are so many preference factors that obtaining data on all of them is not always easy. They range from simple factors such as age, gender, education, and income, to complex and specific factors such as sexual orientation, music taste, and social values [15], [16]. Customer segmentation enhances understanding of customer needs, preferences, and behavior. The implementation of segmentation enables Fast-Moving Consumer Goods (FMCG) stores to provide more personalized and relevant customer services, support purchasing behavior analysis, and build strong relationships with consumers, leading to sustainable business growth [17], [18].

This research utilizes customer segmentation based on purchasing potential in FMCG stores. FMCG stores are retail outlets that focus on selling consumer products with high turnover rates. These products are daily necessities for consumers and have fast consumption characteristics. FMCG stores typically offer affordable products and are often strategically located for easy customer accessibility [19]. Products sold in FMCG stores include various categories such as food and beverages, personal care and hygiene products, health items, household essentials, and children and baby products. Analyzing purchasing potential can aid in identifying market opportunities, developing effective pricing strategies, targeting promotions, managing inventory and distribution, and improving products to enhance customer satisfaction. The most commonly used segmentation method employs the K-Means Clustering algorithm, a machine learning algorithm for grouping data into homogeneous clusters [20].

Research on online sales customer segmentation using the K-Means clustering method varies. One study used transaction data from online/retail stores in the United Kingdom registered with an online retail company from December 1, 2010, to December 9, 2011, totaling 541,909 transactions. The results revealed two clusters: cluster 0, with customers having higher total transactions, and cluster 1, with customers making the most frequent transactions [2]. Another study focused on customer segmentation using the K-Means method with data from 9,000 credit card users in a bank. The results identified three clusters: cluster 1 with moderate credit card usage, cluster 2 with the least credit card usage, and cluster 3 with the most frequent credit card usage and product purchases [21]. A different study applied the K-Means method to customer grouping in UMS store. This research successfully created a customer grouping system with three clusters based on a sample of 357 data. Cluster 3, with 10 data points at the final centroid, was identified as the potential customer group, recommended for UMS Store [22]. A different study applied the K-Means method to customer grouping in the UMS store. This research successfully created a customer grouping system with three clusters based on a sample of 357 data points. Cluster 3, with 10 data points at the final centroid, was identified as the potential customer group recommended for the UMS store [22].

A study on customer segmentation based on loyalty levels used K-Means and LREM feature selection in an online retail store. This research utilized a dataset from the Kaggle online retail platform between January 4, 2011, and December 10, 2011, consisting of 37,980 records with 8 different attributes. The segmentation divided customers into four groups: Premium Loyalty, Inertia Loyalty, Latent Loyalty, and Not Loyal [23]. Multichannel marketing customer segmentation research helps in comparing multi-dimensional consumer buying patterns with interconnected variables [24].

K-Means clustering method has been applied in various research and business contexts. For instance, in the studies mentioned, the method was used to group customers in online sales based on factors such as purchasing behavior, price trends, and demographic data. The results indicated that the application of customer segmentation can assist e-commerce companies in designing more effective marketing strategies [21].

By applying this method to customer data, companies can identify potential segments and optimize product offerings, pricing, and services. Based on previous research testing the success of the K-Means method in various contexts, the objective of this research is customer segmentation based on purchasing potential using the K-Means Clustering method. This study can provide valuable insights for companies

TABLE 1
DATASET DESCRIPTION

Attribute	Description
Id	Display customer code from 1-2000
Sex	Gender: 0= Male, 1 = Female
Marital Status	Marital status: 0= Single, 1=Married
Age	Age (years): minimum= 18, maximum= 78
Education	Level of customer's education: 0= other/unknown, 1= high school, 2= university, 3= graduate school
Income	Annual Revenue (US$): Minimum=$35832, Maximum= $309364
Occupation	Occupation: 0= Unemployed/Unskilled, 1= Skilled employee/Official, 2= Management/Self-employed/High-quality employee/Official.
Settlement Size	Settlement size of the city where the customer is: 0= Small town, 1= Medium city, 2= Large city

TABLE 2
ROW DATASET BEFORE NORMALIZATION

Id	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
100000001	0	0	67	2	124670	1	2
100000002	1	1	22	1	150773	1	2
100000003	0	0	49	1	89210	0	0
100000004	0	0	45	1	171565	1	1
100000005	0	0	53	1	149031	1	1
100000006	0	0	35	1	144848	0	0
100000007	0	0	53	1	156495	1	1
…..	…..	…..	…..	……..	…..	…..	…..
100001998	0	0	31	0	86400	0	0
100001999	1	1	24	1	97968	0	0
100002000	0	0	25	0	68416	0	0

in improving customer experience, increasing customer retention, and enhancing marketing efficiency through a better understanding of customer needs and behavior.

II. RESEARCH METHOD

This research involves a series of important steps to prepare and plan the research stages comprehensively. A literature review is a thorough examination of literature and information sources relevant to a research topic [8], [9]. It aims to identify knowledge gaps, build a conceptual framework, evaluate research methods, provide historical context, and support arguments and analyses in the research. The subsequent stages conducted in this research are outlined as follows:

A. Data collection

This research collects data by searching for customer segmentation datasets on the Kaggle website by Dev Sharma as the research object [10]. The dataset description is presented in Table 1. Source link: [Kaggle Dataset].

B. Customer Segmentation Dataset

The data collected consists of customer data. The row dataset for customer segmentation is presented in Table 2.

C. Data Pre-processing

The data pre-processing stage involves checks for missing values and data normalization [25], [26]. Checking for missing values involves identifying and handling empty data in the dataset, while data normalization scales the values of variables uniformly. Both steps are crucial in the implementation of k-means clustering, where checking for missing values ensures data integrity before analysis, and normalization helps address scale issues that can affect clustering results [27], [28]. Normalization uses feature scaling technique as shown in (1). Here, X_{new} denotes the new value after normalization, X_{old} is the initial data value before normalization, and X_{max} is the maximum value among all data in the same attribute [29].

$$X_{new} = \frac{X_{old}}{X_{max}} \quad (1)$$

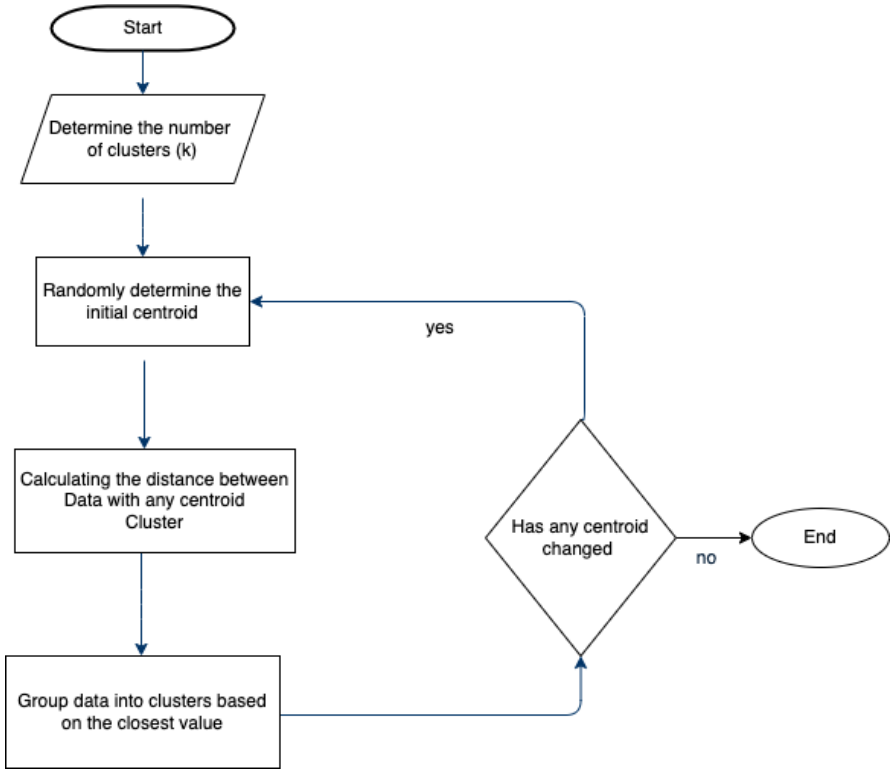

Figure 1. The flowchart of K-Means Clustering

D. Application of K-Means Clustering

The next step after preprocessing is the implementation of K-Means clustering. K-Means clustering, introduced by Macqueen, is the most famous algorithm in clustering techniques [30]. It is also one of the most prominent clustering techniques in science and technology [31]. Its goal is to divide data into several groups [32]. The flowchart of K-Means Clustering can be seen in Figure 1.

The following is the explanation of the stages of the K-Means Clustering Flowchart in Figure 1 [33]. First, determining the number of clusters (*k*) is the first step in finding the optimal number of clusters for the data to be grouped. Second, determining the initial centroids randomly in each cluster. The initial centroid selection is done by randomly selecting data points from the dataset. The data points chosen to be initial centroids are selected without considering the distribution or specific characteristics of the data. Each point in the dataset has an equal chance of being chosen as an initial centroid. Third, calculating the distance between each data point and every cluster centroid. In K-Means Clustering, the distance between each data point and every cluster centroid is calculated using the Euclidean Distance formula. To calculate the distance between data point *x* row *I* (*i*=1,2,3,..,*n*), and data point *c* row *h* (*h*=1,2,3,…,*k*), denoted as $d\ (x_i, c_h)$, where *n* is the total number of data rows, *m* is the number of attributes, and *k* is the number of clusters. The distance formula $d\ (x_i, c_h)$is shown in (2).

$$d\ (x_i, c_h) = \sqrt{\sum_{j=1}^{m}\sum_{h=1}^{k}(x_{ij} - c_{hj})^2} \tag{2}$$

$$d\ (x_i, c_h) = \sqrt{(x_{i1} - c_{h1})^2 + (x_{i2} - c_{h2})^2 + (x_{i3} - c_{h3})^2 + \ldots + (x_{im} - c_{hm})^2}$$

Where x_{ij} is the *j* attribute of data *i* and c_{hj} is the *j* attribute of cluster *h*. The smallest distance from data *i* to cluster *h* indicates that data *i* belongs to cluster *h*. If the distance from data 5 is smallest with cluster 3, then data 5 is grouped into cluster 3. After determining the cluster of each data point, the average value of each attribute in each cluster is calculated to generate new centroids for those clusters.

TABLE 3
ROW DATASET AFTER NORMALIZATION

Id	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
100000001	0	0	0.882	0.667	0.403	0.5	1
100000002	1	1	0.289	0.333	0.487	0.5	1
100000003	0	0	0.645	0.333	0.288	0	0
100000004	0	0	0.592	0.333	0.555	0.5	0.5
100000005	0	0	0.697	0.333	0.482	0.5	0.5
100000006	0	0	0.461	0.333	0.468	0	0
100000007	0	0	0.697	0.333	0.506	0.5	0.5
…..	…..	…..	…..	……	…..	…..	…..
100001998	0	0	0.408	0	0.279	0	0
100001999	1	1	0.316	0.333	0.317	0	0
100002000	0	0	0.329	0	0.221	0	0

TABLE 4
EARLY CENTROID

Centroid	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
k1	0	0	0.645	0.333	0.288	0	0
k2	0	0	0.697	0.333	0.506	0.5	0.5
k3	1	1	0.329	0.333	0.351	0	0

The final stage is to check whether there is a change from the new centroids compared to the previous centroids after grouping the data into clusters. If there is a change in centroid values, the process is still ongoing, and data grouping must continue in the next iteration. Repeat steps 2, 3, and 4 in the next iteration until there are no further changes in the centroid of each cluster. If there are no changes in the centroids, the clustering process is considered complete.

E. Cluster Results Analysis

In cluster analysis, customer segmentation groups are selected based on purchasing potential, categorized as high, medium, and low.

III. RESULT AND DISCUSSION

A. Pre-processing

The pre-processing analysis stage, which includes handling missing values and data normalization, is shown in Table 3.

B. K-Means Clustering

Determining the number of clusters is based on attributes that are significant factors influencing customer purchasing power [20]. Customers with more financial resources have higher purchasing power to meet consumption needs and desires, including purchasing goods and services. High-income levels are often associated with a higher quality lifestyle. Customers with high incomes tend to spend more on experiences, travel, or luxury products. This study categorizes customers into three clusters: cluster 1 ($k1$) represents customers with low purchasing potential, cluster 2 ($k2$) represents customers with moderate purchasing potential, and cluster 3 ($k3$) represents customers with high purchasing potential.

This stage is the first iteration, with the determination of initial centroids done randomly. Initial centroids are the initial values of the cluster centers that will be identified and updated during the algorithm iterations to achieve the final configuration of the clusters. The cluster centers or initial centroids are presented in Table 4.

The distance between the first data point and the initial centroid points (Table 4) is calculated using (1). This involves calculating $d(1,1)$, the distance between the first data point and the first centroid of cluster 1 as 0; $d(1,2)$, the distance between the first data point and the second centroid of cluster 2 as 1; and $d(1,3)$, the distance between the first data point and the third centroid of cluster 3 as 1.

The calculation of distances (d) from the first data point to the initial centroids, consisting of 7 attributes, results in grouping into cluster 1, because the distance from the first data point to the first centroid of cluster 1, $d(1,1)$, is the smallest, with a value of 0. The distances to the second centroid of cluster 2, $d(1,2)$, and the third centroid of cluster 3, $d(1,3)$, are both 1. The process for calculating

TABLE 5
TABLE 8
CENTROID ITERATION RESULT 4

Centroid	Sex	Marital status	Age	Education	Income	Occupation	Settlement size	Number of Data
k1	0.263	0.103	0.482	0.273	0.314	0.125	0.020	535
k2	0.003	0.238	0.510	0.315	0.464	0.628	0.751	685
k3	0.947	0.942	0.435	0.415	0.376	0.385	0.260	780

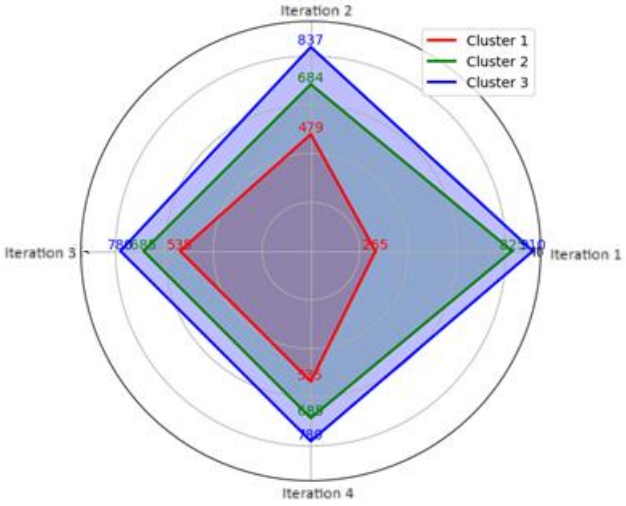

Figure 2. Radar Plot Results of the Sum of Cluster Data for Each Iteration

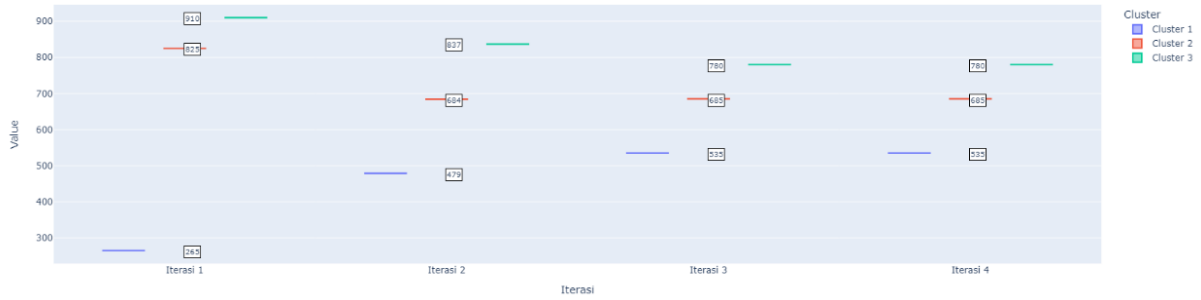

Figure 4 Box Plot Results of the Sum of Cluster Data for Each Iteration

distances from the second to the last data points to centroid clusters 1 to 3 follows the same method as the calculation of $d(1,1)$ to $d(1,3)$. The overall distance calculation results in the clustering of the first iteration, directed to Table 5, the clustering results.

Based on the clustering results of iteration 1, the average value of the 6 attributes for each cluster is calculated, as presented in Table 6, resulting in new centroids. Table 6 shows the centroid data of each cluster (k) along with the number of data points. The results show that cluster 1 ($k1$) consists of 265 customers, cluster 2 ($k2$) consists of 825 customers, and cluster 3 ($k3$) consists of 910 customers. For the second iteration, the same process as in the first stage is used, calculating the distance of each data point to the new centroids in Table 6. The K-Means Clustering calculation ends at the fourth iteration, with the clustering results presented in Table 7.

The average calculation of each attribute in iteration 4 results in new centroids, which are displayed in Table 8. In Table 8, the centroid data of each cluster (k) along with the number of data points are shown. The results indicate that cluster 1 ($k1$) consists of 535 customers, cluster 2 ($k2$) consists of 685 customers, and cluster 3 ($k3$) consists of 780 customers. Since the centroid results and the number of data points in clusters for iteration 3 are the same as those in iteration 4, the calculation is terminated and declared complete. The number of data points for each iteration is presented in Figure 3.

In Figures 2 and 3, cluster 2 has the same number of data points from iteration 2 to iteration 4, totaling 685 customers. Meanwhile, cluster 1 experienced a change in the number of customer data points, culminating in iteration 4 with 535 customers, and cluster 3 with 780 customers.

The data interpretation reveals that cluster 1 ($k1$) tends to consist of customers who are male (Sex=0),

unmarried (Marital status=0), with an average age of 37 years, an annual income of around $97,107, high school (Education = 1), mostly unskilled or unemployed (Occupation=0), and residing in small towns (Settlement size=0). The total number of data points in this group is 535. Cluster 2 (*k*2) generally comprises customers who are male (Sex=0), unmarried (Marital status=0), with an average age of 39 years, high school (Education = 1), a high annual income of about $143,402, mostly employed as skilled workers or officials (Occupation=1), and residing in large cities (Settlement size=2). There are 685 data points in this group. Cluster 3 (*k*3) mainly consists of customers who are female (Sex=1), married (Marital status=1), with an average age of 33 years, high school (Education = 1), an annual income of approximately $116,410, mostly employed in management, self-employed, skilled workers, or high-level officials (Occupation=1), and residing in medium-sized cities (Settlement size=1). The total number of data points in this group is 780.

Based on the clustering results of the customer segmentation dataset, it can be divided into three groups based on purchasing potential. Cluster *k*2 indicates high purchasing potential with an annual income of around $143,402, suggesting that customers with higher financial resources may be more inclined to purchase exclusive or higher-priced products. Cluster *k*3, with an income of about $116,410, indicates moderate purchasing potential, where customers have good financial capability. Meanwhile, cluster *k*1, with an income of around $97,107, depicts low purchasing potential, possibly constrained by financial limitations. Tailored marketing strategies adapted to the characteristics of each group can help FMCG stores improve sales effectiveness and better meet customer needs.

It can be interpreted that Cluster *k*2, which has a relatively high income, may have higher purchasing potential because they have more financial resources to make purchases. Customers in cluster *k*2 can be targeted for more exclusive and expensive product offerings or promotions. Cluster *k*1 and *k*3 with lower or moderate income characteristics can be the focus of more affordable and tailored marketing strategies based on their customer profiles.

Research on FMCG store customer segmentation using the k-means clustering method provides in-depth insights into customer characteristics based on purchasing potential. Cluster analysis yields three main groups with different characteristics, allowing stores to understand the needs and preferences of each customer group. The benefits are significant, including the ability to design more effective marketing strategies, improve inventory management, and provide more personalized services.

Suggestions for further research include developing more complex cluster models, validation involving business experts, using additional data such as product preferences, and temporal analysis to understand changes in customer behavior over time. Further research should also focus on developing specific and effective marketing strategies. With these steps, future research is expected to provide stronger guidance for FMCG stores in improving their business performance.

IV. CONCLUSION

The cluster analysis results on the customer segmentation dataset identified three customer groups with different characteristics. Cluster *k*1 is dominated by single male customers with low income, residing in small towns, and tending to have low purchasing potential. Conversely, cluster *k*2, featuring single male customers with high income, residing in large cities, shows higher purchasing potential. Meanwhile, cluster *k*3, with moderate income, encompasses married female customers, residing in medium-sized cities, and having moderate purchasing potential. This analysis illustrates that income significantly influences purchasing potential, where customers with higher income tend to have greater purchasing potential. Tailored marketing strategies adapted to the profile of each group can help FMCG stores improve sales effectiveness and customer satisfaction, with cluster *k*2 being a potential target for exclusive products or promotions. Suggestions for further research involve the development of more complex cluster models and validation involving business experts to enhance the accuracy and relevance of findings.

ACKNOWLEDGMENT

Thank you to Kaggle for providing the valuable dataset and to the Master Program of Informatics Universitas Ahmad Dahlan for the essential support in this research. Collaboration and support from all

parties are highly appreciated for the progress of this research. Hopefully, the results contribute positively to the advancement of science and public health.

REFERENCES

[1] R. Daraboina, O. Cooper, and M. Amini, "Segmentation of organic food consumers: A revelation of purchase factors in organic food markets," *J. Retail. Consum. Serv.*, vol. 78, p. 103710, 2024.
[2] C. Hafidz Ardana, A. A. A. A. Khoyum, and M. Faisal, "Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024.
[3] R. A. Utamanyu and R. Darmastuti, "Budaya Belanja Online Generasi Z dan Generasi Milenial di Jawa Tengah (Studi Kasus Produk Kecantikan di Online Shop Beauty by ASAME)," *Scriptura*, vol. 12, no. 1, pp. 58–71, 2022.
[4] L. Safitri and F. Fathurohman, "Keputusan Pembelian Pangan Segar secara Online pada Masa Pandemi Covid-19: Fresh Food Online Purchase Decision During The Covid-19 Pandemic," *J. Agroindustri Terap. Indones.*, vol. 1, no. 2, pp. 1–10, 2024.
[5] T. K. H. Chan, C. M. K. Cheung, and Z. W. Y. Lee, "The state of online impulse-buying research: A literature analysis," *Inf. Manag.*, vol. 54, no. 2, pp. 204–217, 2017.
[6] R. Daulay and S. Handayani, "Pengaruh Gaya Hidup dan Kepribadian terhadap Minat Belanja Online melalui Aplikasi Lazada," in *Seminar Nasional Teknologi Edukasi Sosial dan Humaniora*, 2021, pp. 478–486.
[7] D. Suleman, "Faktor penentu keputusan konsumen Indonesia memilih tempat belanja disebuah e-commerce (Theory of Planned Behavior)," *J. Dr. Manaj.*, vol. 1, pp. 1–9, 2018.
[8] M. Skare, B. Gavurova, and M. Rigelsky, "Innovation activity and the outcomes of B2C, B2B, and B2G E-Commerce in EU countries," *J. Bus. Res.*, vol. 163, p. 113874, 2023.
[9] A.-Y. Al-Yasir, M. Afdal, Z. Zarnelly, and A. Marsal, "Analisis Loyalitas Pelanggan Business To Business Berdasarkan Model RFM Menggunakan Algoritma Fuzzy C-Means: Business to Business Customer Loyalty Analysis Based on RFM Model Using Fuzzy C-Means Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 359–365, 2024.
[10] D. E. B. Jabat, L. L. Tarigan, M. Purba, and M. Purba, "Pemanfaatan Platform E-Commerce Melalui Marketplace sebagai Upaya Peningkatan Penjualan," *Skylandsea Prof. J. Ekon. Bisnis Dan Teknol.*, vol. 2, no. 2, pp. 16–21, 2022.
[11] A. Asyifah, A. Syafi'i, H. Hanipah, and S. Ispiyani, "Pengembangan Aplikasi E-Commerce Untuk Peningkatan Penjualan Online," *Action Res. Lit.*, vol. 7, no. 1, pp. 70–75, 2023.
[12] M. R. Olii, M. H. Pratiknjo, and S. Rumampuk, "Online Shop Sebagai Alternatif Berbelanja Masyarakat Kota Manado," *HOLISTIK, J. Soc. Cult.*, 2020.
[13] R. A. Sulistianti and N. Sugiarta, "Konstruksi Sosial Konsumen Online Shop Di Media Sosial Tiktok (Studi Fenomenologi Tentang Konstruksi Sosial Konsumen Generasi Z Pada Online Shop Smilegoddess Di Media Sosial Tiktok)," *JISIP (Jurnal Ilmu Sos. dan Pendidikan)*, vol. 6, no. 1, 2022.
[14] D. M. Mansur, E. T. Sule, D. Kartini, Y. M. Oesman, A. H. P. K. Putra, and N. Chamidah, "Moderating of the role of technology theory to the existence of consumer behavior on e-commerce," *J. Distrib. Sci.*, vol. 17, no. 7, pp. 15–25, 2019.
[15] F. Sudirjo, T. Purwati, W. Widyastuti, Y. U. Budiman, and M. Manuhutu, "Analisis Dampak Strategi Pemasaran Digital dalam Meningkatkan Loyalitas Pelanggan: Perspektif Industri E-commerce," *J. Pendidik. Tambusai*, vol. 7, no. 2, pp. 7524–7532, 2023.
[16] İ. Sabuncu, E. Turkan, and H. Polat, "Customer segmentation and profiling with RFM analysis," *Turkish J. Mark.*, vol. 5, no. 1, p. 22, 2020.
[17] J. Anselmsson and N. Bondesson, "Brand value chain in practise; the relationship between mindset and market performance metrics: A study of the Swedish market for FMCG," *J. Retail. Consum. Serv.*, vol. 25, pp. 58–70, 2015.
[18] F. Dweiri, S. A. Khan, and A. Almulla, "A multi-criteria decision support system to rank sustainable desalination plant location criteria," *Desalination*, vol. 444, pp. 26–34, 2018.
[19] G. Pierce and R. Connolly, "Disparities in the 'who' and 'where' of the vehicle purchase decision-making process for lower-income households," *Travel Behav. Soc.*, vol. 31, pp. 363–373, 2023.
[20] S. Bogomolova, Z. Anesbury, L. Lockshin, N. Kapulski, and T. Bogomolov, "Exploring the incidence and antecedents of buying an FMCG brand and UPC for the first time," *J. Retail. Consum. Serv.*, vol. 46, pp. 121–129, 2019.
[21] F. D. S. Alhamdani, A. A. Dianti, and Y. Azhar, "Segmentasi Pelanggan Berdasarkan Perilaku Penggunaan Kartu Kredit Menggunakan Metode K-Means Clustering," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 2, pp. 70–77, 2021.
[22] M. A. Ghofar and Y. I. Kurniawan, "Aplikasi Pengelompokan Pelanggan Pada Ums Store Menggunakan Algoritma K-Means," *J. Teknol. dan Manaj. Inform.*, vol. 4, no. 1, 2018.
[23] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus: Belle Crown Malang)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput. E-ISSN*, vol. 2548, 2009.
[24] H.-H. Zhao, X.-C. Luo, R. Ma, and X. Lu, "An extended regularized K-means clustering approach for high-dimensional customer segmentation with correlated variables," *Ieee Access*, vol. 9, pp. 48405–48412, 2021.
[25] N. N. Almanza-Ortega *et al.*, "Corrosion Analysis Through an Adaptive Preprocessing Strategy Using The K-Means Algorithm," *Procedia Comput. Sci.*, vol. 219, pp. 586–595, 2023.
[26] F. Barrera, M. Segura, and C. Maroto, "Multiple criteria decision support system for customer segmentation using a sorting outranking method," *Expert Syst. Appl.*, vol. 238, p. 122310, 2024.
[27] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, "An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity," *Decis. Anal. J.*, vol. 9, p. 100341, 2023.
[28] H. Ghaderi, B. Foreman, A. Nayebi, S. Tipirneni, C. K. Reddy, and V. Subbian, "A self-supervised learning-based approach to clustering multivariate time-series data with missing values (SLAC-Time): An application to TBI phenotyping," *J. Biomed. Inform.*, vol. 143, p. 104401, 2023.
[29] S. H. Nam, S. Kwon, and Y. Do Kim, "Development of a basin-scale total nitrogen prediction model by integrating clustering and regression methods," *Sci. Total Environ.*, p. 170765, 2024.
[30] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, "Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm," *Appl. Soft Comput.*, vol. 113, p. 107924, 2021.
[31] M. Faisal and E. M. Zamzami, "Comparative analysis of inter-centroid K-Means performance using euclidean distance, canberra distance and manhattan distance," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 12112.
[32] K. Sya'iyah, H. Yuliansyah, and I. Arfiani, "Clustering student data based on K-means algorithms," *Int. J. Sci. Technol. Res*, vol. 8, no. 8, pp. 1014–1018, 2019.
[33] Q. Liu, J. Ma, X. Zhao, K. Zhang, K. Xiangli, and D. Meng, "A novel method for fault diagnosis and type identification of cell voltage inconsistency in electric vehicles using weighted Euclidean distance evaluation and statistical analysis," *Energy*, p. 130575, 2024.