



Customer Clustering Using The K-Means Clustering Algorithm in Shopping Mall in Indonesia

Iksan Adityo Mulyo^{1✉}, Jerry Heikal²

¹Management, Central Queensland (CQU) - Bakrie University, Indonesia

²Business Administration, Central Queensland (CQU) - Bakrie University, Indonesia

Article Information

Article History:

Received November 2022

Revised November 2022

Accepted December 2022

Keywords:

Shopping mall, data mining, customer clusterization, clustering, k-means

Abstract

Shopping Mall in Indonesia offers several kinds of product for customer to buy. Determining client interest in purchasing a product is difficult because the number of visitors varies. As a marketing plan strategy, it is supposed to market the new product similar to customers' interests. This study aims to identify the customer's characteristics, product categories and target segmentation in the shopping mall using the K-Means Clustering algorithm. There were ten variables, age, sex, annual income, spending score, marital status, home location, work location, product interest, social media and children. The application of SPSS V.25 assisted data processing. The total population in this study were 200 participants. In this case, there were 3 clusters with each persona. First Cluster was Gadget Lovers, the second Cluster was Beauty Influencer, and 3rd Cluster was Mature Beauty. The target segment was the customer with profile men & women ages 29 and 40 years old and annual incomes IDR 65,000,000 and 87,000,000, with Instagram as their favorite social media. The product that should be more focused on was Gadget (hand phone & Tablet + accessories) since 1st Cluster (Gadget Lovers) & 2nd the Cluster (Beauty Influencers), who had a total of 58% of the total population, was likely to buy that product in Shopping Mall. Based on the author's buyer persona, Online Valuable Proposition (OVP) for this product is Trusted.

INTRODUCTION

Shopping Mall in Indonesia provides several kinds of product that customer can purchase. The number of customer visits fluctuates, make it difficult to identify the customer's interest in purchasing a product. The 2020 IBM US Retail Index Report stated that the Covid-19 pandemic accelerated the change in shopping behavior to online activities about 5 years earlier. This change is a necessity. The reason is that the Covid-19 pandemic has prompted the governments of various countries, including Indonesia, to limit the movement of people to prevent the spread of the virus, including restrictions on visits to shopping cen-

ters or malls (Sihombing & Nasib, 2020). Mandiri Institute report in 2021 reported that the number of visits to shopping places continued to increase until early September 2021 (Jayani, 2021).

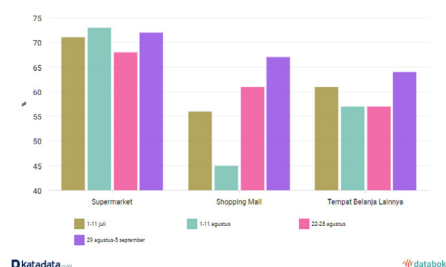


Figure 1. Visits Frequency to Shopping Mall in 2021

✉ Correspondence Address:

Jl. H. R. Rasuna Said, Setiabudi, Kuningan,

Daerah Khusus Ibukota Jakarta 12940

E-mail: iksan.adityo17@gmail.com,

jerry.heikal@bakrie.ac.id

Based on the picture above, it can be seen that visits to shopping malls have continued to increase since early August 2021. There were 45% of visits to shopping malls during peak hours on August 1 to 11, 2021. This visit increased to 61% on August 22 to 28 2021, and again to 67% on August 29 to September 5 2021. Visits to supermarkets also increased after experiencing a decline from August 1 to 11, 2021. This increase was from 68% on August 22 to 28, 2021 to 72% on August 29 to September 5, 2021 during peak hours.

Mandiri Institute also said that visits to other shopping places increased from 57% on August 22 to 28, 2021 to 64% on August 29 to September 5, 2021. There is a difference in the recovery in the number of visits. The report explained that there is a difference between shopping malls and supermarkets. Shopping malls are generally shopping centers that offer a variety of consumer experiences. Meanwhile, supermarkets are more focused on daily household needs.

On the other hand, Katadata Insight Center researched the behavior of customers in Indonesia. The research reached more than 10 million transaction samples. The data comes from nearly one million credit card users from January to December 2020.

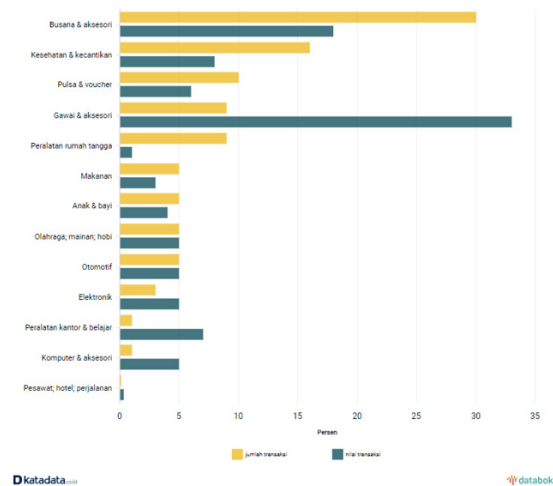


Figure 2. Customers' Behavior based on the product bought in Shopping Mall

The picture above shows that fashion or clothing products and their accessories are the most purchased items by consumers. These products dominate at least 30% of transactions. The transaction value of fashion and accessory products is only 18%. Today's society wants more diverse products that follow the latest trends, so that fashion product is sought after by consumers. In addition, these products can also be purchased at

affordable prices. Meanwhile, more than a quarter of the transaction value was controlled by gadgets and accessories. The value reaches 33%, although the number of transactions is only 9% because the average value of the device per transaction reaches IDR 692,481.

Based on the data shown above, there is an opportunity to enhance and enlarge the market of a particular product in the shopping mall. Before doing digital market strategy, ensure every stakeholder involved knows how to manage and maintain the customer relationship between seller and customers. Managing and maintaining customer relationships have always been critical in providing business intelligence to organizations to build, manage and develop valuable long-term customer relationships. The importance of treating customers as an organization's primary asset is increasing in value in the present day. Organizations are interested in investing to the development of customer acquisition, maintenance and strategies. Business intelligence is vital in allowing companies to use technical expertise to gain better customer knowledge. In programs for outreach using clustering techniques like k-means, customers with similar means are clustered together (Dey, 2019). Customer segmentation helps the marketing team to identify and expose different customer segments that have unique thought and follow different purchasing strategies. Customer segmentation helps determine the customers who vary in preferences, expectations, desires and attributes. The primary purpose is to group people with similar interests thus the marketing team can converge on an effective marketing plan. Clustering is an iterative process of knowledge discovery from vast amounts of raw and unorganized data. Clustering is a form of exploratory data mining used in many applications, including machine learning, classification and pattern recognition (Dedhia, 2020).

The novelty of this research is gives the precise profiling of target segment, such as their income, their ages, and their favorite of social media, then producer can define the persona of each buyer. So, it will give the producer what kind of product that should be focused on to sell. It will optimizing marketing activity in order to maximize profit.

Based on the explanation above, this study aims to identify the customer's characteristics, product categories and target segmentation in the shopping mall using the K- Means Clustering algorithm. As a marketing plan strategy, it is supposed to market the new product similar to customers' interests.

Over the years, as there has been extreme competition in the business world, organizations have to enhance their profits by satisfying their customers' demands and attracting new customers according to their needs. Identifying customers and meeting each customer's demands was a very complex task. Customers may differ according to their demands, desires, preferences, etc. Instead of a "one-size-fits-all" approach, customer segmentation clusters the customers into groups sharing the same properties or behavioral characteristics. Customer segmentation was a strategy of dividing the market into homogeneous groups (Kushwaha & Prajapati, 2020). The customer segmentation technique allowed the business to use its marketing budgets better and gain a competitive edge over rival companies, demonstrating a better knowledge of customer needs. It also helps an organization increase its marketing efficiency, plan its budget, determine new market opportunities, make better brand strategies, and identify customer retention. The data employed in the customer segmentation technique, that divides the customers into groups, is influenced by a variety of elements such as demographical and geographical data conditions, economic conditions, and behavioral patterns. Further explanation of each factor can be seen in the following:

Demographic variables such as age, gender, family, education level, and income were the segmentation's most accessible and standard variables (Nica et al., 2019). Demographical segmentation can also help business owners to determine their marketing strategy. For example, if a company targets teenagers, it will use a digital marketing strategy. Meanwhile, they will most likely use a conventional marketing strategy if they target middle-aged customers. This method is one of the most used methods to identify and analyze customer behavior and segmentation.

This method divided the customer segmentation based on a particular location. Small business owners mainly utilize geographic segmentation because this method is objective, relatively easy, and cheaper than demographic, economic, or behavioral segmentations. Usually, business owners target people who live near their shops or assume that a group of people who live in the same area have similar needs (Musyoka et al., 2007).

Economic segmentation is derived from the class difference in society. This method assumes that people with different economic statuses will have different needs and interests (Dzobo et al., 2014). For example, people in the lower eco-

nomic class will focus on fulfilling their basic needs, namely food, clothes, and shelter. Meanwhile, people in the higher economic class do not have to worry about it since they have no problem fulfilling their primary and secondary needs. Therefore, they can spend their wealth on tertiary needs. Using economic conditions to determine customer segmentation, business owners can also determine the appropriate price rates to sell their products and services.

This method divides customers based on their interactions with certain brands, websites, knowledge, use, and purchasing tendencies (Bang et al., 2015). These behavioral patterns are later analyzed to provide a complete understanding for business owners. This way, business owners can curate a unique promo collection for specific customers and determine the most suitable marketing campaigns.

METHOD

The data in this study is taken from Kaggle datasets. There are 10 variables: age, gender, annual income, spending score, marital status, home location, work location, product interest, social media and children. The total population in this study are 200 participants.

In order to understand the customer, especially the target customer, the sensor is provided to the marketing team thus the strategy can be planned accordingly. Spending Score is assigned to the customer based on defined parameters like customer behavior and purchasing data (Tan et al., 2005).

The research method using K-Means Clustering, this study uses 10 variables to segment the shopping mall customers in Indonesia. K-means is one of the most popular clustering algorithms, mainly because of its good time performance (Dey, 2019).

K-means algorithms is an algorithms which needs an input parameter "k" and separated object "n" to cluster "k" so that the similarity level between member in one cluster is high meanwhile the similarity level in other cluster is low. The similarity of members to the cluster is measured by the proximity of the object to the mean value in the cluster or can be referred to as the centroid cluster.

K-means algorithms is to determine the cluster total (K), set any cluster center, and then calculate the distance of each data to the center of the cluster, and classify the data to the cluster with the shortest distance, calculate the cluster center, then

Process clustering start to identify a data which will be clustered, X_{ij} ($i=1, \dots, n$; $j=1, \dots, m$) with n is a total data which will be clustered and m is a total of variable. At the beginning of iteration, the center of each cluster was determined by free, C_{kj} ($k=1, \dots, K$; $j=1, \dots, m$). Then calculated the distance between data and cluster center.

To calculate the distance of data to -i (X_i) to the cluster center to -k (C_k), name (d_{ik}), can be used in Euclidean formulas as below:

$$d_{ij} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

A data become a member from cluster to-J if the data distance to the cluster center to-J which lower value than the distance to other cluster center.

The new cluster center value can be calculated by finding the average value from the data which become a member to those cluster with formulas:

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{hj} \in \text{cluster } k - k \quad (2)$$

Based on the calculating the data with K-means algorithms method, the research also use the application of SPSS V.25 to assists in the data processing. After the data is processed, there are 3 clusters with every persona we have, 1st Cluster is Gadget Lovers, 2nd Cluster is Beauty Influencer, and 3rd Cluster is Mature Beauty.

Table 1. Variables and Coding Data

AGE, ANNUAL INCOME, AND SPENDING SCORE	
Various numeric	1-200
Tangerang	4
Bandung	5
GENDER	
Female	0
Male	1
Work Location	
Jakarta	1
Bogor	2
Depok	3
Tangerang	4
Bekasi	5
SOCIAL MEDIA	
YouTube	0
Twitter	1
Facebook	2
Instagram	3

WhatsApp	4
TikTok	5
Line	6
Tumblr	7
Pinterest	8
Telegram	9
Reddit	10
Snapchat	11
LinkedIn	12
CHILDREN	
1-2 Child	0
3-4 Child	1
5-more child	2
AGE, ANNUAL INCOME, AND SPENDING SCORE	
Various numeric	1-200
Tangerang	4
Bekasi	5
GENDER	
Female	0
Male	1
Work Location	
Jakarta	1
Bogor	2
Depok	3
Tangerang	4
Bekasi	5
SOCIAL MEDIA	
YouTube	0
Twitter	1
Facebook	2
Instagram	3
WhatsApp	4
TikTok	5
Line	6
Tumblr	7
Pinterest	8
Telegram	9
Reddit	10
Snapchat	11
LinkedIn	12
CHILDREN	
1-2 Child	0
3-4 Child	1
5-more child	2

RESULT AND DISCUSSION

Using K-Means Clustering, this study uses 10 variables to segment the shopping mall customers in Indonesia. The ten variables are age,

gender, annual income, spending score, marital status, home location, work location, product interest, social media and children. The total population in this study are 200 participants. Before running the K-Means Cluster test, the research data was turned into a "dummy variable" to make customers dividing easier (Danurisa & Heikal, 2022).

We can see the final cluster centers. The results are summarized in Table 2, i.e. which cluster each unit belongs to and the new cluster centers. In the table 2 below, the data set is already divided into 3 clusters according to the data set variables.

Table 2. Final Cluster Centers

	Final Cluster Centers		
	Cluster		
	1	2	3
Gender Code	0.53	0.45	0.39
Age	40	29	47
Annual Income (in Million Rp)	87	65	45
Spending Score	19	75	42
Marital Status	0.84	0.40	0.86
Home Location	3	3	3
Work Location	3	3	3
Product Interest	8	8	9
Social Media	3	3	3
Children	1.16	0.42	1.32

Cluster No.1 has 19% of the total population, and the customer sex is a man named Harun, age 40. Married status with 1-2 children and domiciled in Depok with an annual income of IDR 87,000,000. His favorite product is a gadget (hand phone & Tablet + accessories). His favorite social media is Instagram. The buyer Persona for cluster 1 is named "The Gadget Lovers".



Figure 3. Gadget Lovers

Cluster No.2 has 39% of the total population, and the customer sex is a woman named Dwi, age 29 years old. Single status and domiciled in Depok with an annual income of IDR 65,000,000. She has a spending score 75. Her favorite product is a gadget (hand phone & Tablet

+ accessories), and her favorite social media is Instagram. The buyer Persona for cluster 2 is named "Beauty Influencer".



Figure 4. Beauty Influencer

Cluster No.3 has a total of 42% of the total population, and the customer sex is a woman named Chintami, age 47. Married status with 1-2 children and domiciled in Depok with an annual income of IDR 45,000,000. Her favorite product to buy is beauty cosmetics. She has a spending score 42. Her favorite social media is Instagram. The buyer Persona for cluster 3 is named "Mature Beauty".



Figure 5. Mature Beauty

More marketing should be targeted at customer profiles, men & women with ages 29 and 40 years old and annual incomes IDR 65,000,000 and 87,000,000, with Instagram as their favorite social media. In this case, the product that should be more focusing is a Gadget (hand phone & Tablet + accessories). since the first cluster (Gadget Lovers) & second cluster (Beauty Influencers), who has a total 58% of the total population is likely to buy that product in Shopping Mall.

Having a value proposition is especially important for online businesses. The author put "Trusted" as an Online Value Proposition (OVP) in this case. To maintain quality after-sales service, the store provides an authorized Service Center for various products and brands of gadgets. The store also receives warranty claims and repairs smartphones, tablets, and accessories from various other products. Besides that, the store provides Return Products or Refunds if any factory manufacturing defects and malfunctions occur.

Content, Customization, and Communi-

ty. Provide various types of gadgets (hand phone & Tablet + accessories) from each brand. Thus, it can attract loyal customers from each brand, put the best shoot of the products on the website, and engage with customers by asking for their comments & testimonies through WhatsApp or e-mail.

Cost Reduction. Offering differential prices – typically lower online and give the customers cashback promotions on certain products. Refund if there is a damaged or unsatisfied product. Simplify the Process. Integrating between offline stores and online stores. This gives consumers a choice of whether to come directly to the store or need go online. Then, provide an Online website, a new place and a new channel.

In addition to providing new methods for customer acquisition through search engines, affiliates, and e-mail blasting, the web is at its best when unique web offers are available in competitions and sales promotions. Moreover, enhanced social media promotion is based on Clustering results by making Social Media (Instagram) accounts in order to reach customers & going viral and use influencers and an active account to post every event in this store.

In this case, tactics Start, Keep & Stop should be implemented in the marketing strategy since we have a scarcity of capital.

Start : E-mail blasting by offering discounts & any promotions (cashback, newest products, etc); Using influencer marketing to attract customers, especially on Instagram; Automated customer service or live chat helps engage with a customer and personalize their experience; Interaction using Chat Bot with customers.

Keep : Proactive customer service with regular check-ins; Social Media Marketing focusing on Instagram; Launching a new product collection every month from various brands; Reaching out to customers to get their reviews & testimonies after they receive their order through WhatsApp or e-mail.

Stop : SMS Marketing; Cold Calling customers; Spread the brochure & newsletter; Delay in complaint resolution; Too much content, and not focus on promotion & sales marketing.

Research conducted by (Dista & Abdulloh, 2022) revealed that they divided mall customers into 5 clusters based on the annual income they get and spending scores. The use of the Particle Swarm Optimization method for cluster optimization produced by the K-Means method proved to be right after being evaluated and succeeded in increasing the cluster results from the K-Means method as evidenced by the Davies Bouldin

Index method. The data that has been obtained does not rule out the possibility that it will change over time. Meanwhile this research analyze regarding marketing strategy through K-Means clustering algorithm method into 3 clusters based on the 10 variables are gender code, age, annual income, spending score, marital status, home location, work location, product interest, social media, children.

DISCUSSION

After the authoring process and analyzing the customer's clustering and persona, the author gets summaries as stated above. There are men & women are likely to buy Gadget & Tablet + accessories at Shopping Mall, especially single women, age 29 years old, with annual income IDR 65,000,000. A single woman with an annual income IDR 65,000,000 has the highest spending shopping at the shopping mall. Married woman, age 47 years old with 1-2 children, who have an annual income IDR 45,000,000 is likely to buy beauty cosmetics, married women & men, also single women, ages 29, 40 and 47 years old, choose Instagram as their favorite social media, there is a product that can develop, which is Beauty Cosmetics, since the persona "Mature Beauty", who has a total 42% of the total population, is likely to buy that product, in this case, the product that should be more focusing is Gadget (hand phone & Tablet + accessories) since the first cluster (Gadget Lovers) & second cluster (Beauty Influencer), who has a total 58% of the total population is likely to buy that product, more marketing activity on Gadget products (handphone & Tablet + accessories) should target customer profile men & women between ages 29 and 40. It should get annual income IDR 65,000,000 and 87,000,000 because, based on the 1st & 2nd Cluster, they have 58% combined with the total population

CONCLUSION

Shopping Mall in Indonesia provides several kinds of product that customer can purchase. The number of customer visits fluctuates, so it is challenging to determine customer interest in purchasing a product. As a marketing plan strategy, it is supposed to market the new product similar to customers' interests. This study aims to identify the customer's characteristics, product categories and target segmentation in the shopping mall using the K-Means Clustering algorithm.

The product that should be more focusing

is Gadget (hand phone & Tablet + accessories) since the first cluster (Gadget Lovers) & second cluster (Beauty Influence r), who has a total of 58% of the total population is, likely to buy that product in Shopping Mall. More marketing should be targeted at customer profile men & women between ages 29 and 40 years old and annual income IDR 65,000,000 and 87,000,000, with Instagram as their favorite social media.

Based on the persona we got, OVP that is suitable for this product is "Trusted". 7P's digital marketing mix should be implemented that matches the Target Segment by using tactics Start, Keep & Stop to achieve the objective for Target Segment, such as 1st Cluster (Gadget Lovers) & 2nd Cluster (Beauty Influencer).

Activities that should start to reach the target segment are e-mail blasting by offering discounts & any promotions (cashback, newest products, etc). Besides that, using influence r marketing to attract customers (especially in Instagram), automated customer service, or live chat helps engage with a customer and personalize their experience and interaction using Chat Bot with customers.

Some activities should be done or continue executing to reach the target segment. Social media marketing focuses on Instagram, launching a new product collection every month from various brands, and reaching out to customers to get their reviews & testimonies after they receive their orders through WhatsApp or e-mail. That activities are proactive customer service with regular check-ins.

However, some activities should stop doing or executing to make marketing effective and efficient. The activities that should stop are SMS Marketing, cold calling customers, spreading the brochure and newsletter, delay in complaint resolution and too much content, and not focusing on promotion and sales marketing.

REFERENCES

- Bang, J., Cho, Y., & Kim, M. S. (2015). Getting Business Insights through Clustering Online Behaviors. *Modelling and Simulation in Engineering*, 2015, 1–8.
- Danurisa, A., & Heikal, J. (2022). Customer Clustering Using the K-Means Clustering Algorithm in the Top 5 Online Marketplaces in Indonesia. *Budapest International Research and Critics Institute-Journal*, 5(3), 24287–24301.
- Dedhia, H. (2020). Hierarchical Clustering for Customer Data. *Kaggle*. Dey, D. (2019). *ML | Mini Batch K-means clustering algorithm*. GeeksforGeeks.
- Dista, T. M., & Abdulloh, F. F. (2022). Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization. *Jurnal Media Informatika Budidarma*, 6(3), 1339.
- Dzobo, O., Alvehag, K., Gaunt, C. T., & Herman, R. (2014). Multi-dimensional customer segmentation model for power system reliability-worth analysis. *International Journal of Electrical Power & Energy Systems*, 62, 532–539.
- Jayani, D. (2021). *Mandiri Institute: Kunjungan ke Tempat Belanja Meningkatkan*.
- Kushwaha, Y., & Prajapati, D. (2020). Customer Segmentation using K-Means Algorithm. *International Journal of Creative Research Thoughts*, 1–4.
- Musyoka, S. M., Mutyauryu, S. M., Kiema, J. B. K., Karanja, F. N., & Siriba, D. N. (2007). Market segmentation using geographic information systems (GIS). *Marketing Intelligence & Planning*, 25(6), 632–642.
- Nica, E., Gajanova, L., & Kicova, E. (2019). Customer segmentation based on psychographic and demographic aspects as a determinant of customer targeting in the online environment. *Littera Scripta*, 12(2), 108–126.
- Sihombing, E. H., & Nasib, N. (2020). The Decision of Choosing Course in the Era of Covid 19 through the Telemarketing Program, Personal Selling and College Image. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 3(4), 2843–2850.
- Tan, P., Steinbach, M., & Kumar, V. (2005). *Basic Concepts and Algorithms. In Introduction to Data Mining, Cluster Analysis*. Addison-Wesley.