Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering

Musthofa Galih Pradana [1,*], Hoang Thi Ha [2],

[1] Department of Informatics Alma Ata University, Yogyakarta, Indonesia
[2] Department Management Information System University of Danang, Vietnam
[1] mgalihpradana@almaata.ac.id; [2] hoang2th@due.udn.vn
* corresponding author

Abstract

The application of customer segmentation is very vital in the world of marketing, a manager in determining a marketing strategy, knowing the target customer is a must, otherwise it will potentially waste resources to pursue the wrong target. Customer segmentation aims to create a relationship with the most profitable customers by designing the most appropriate marketing strategy. Many statistical techniques have been applied to segment the market but very large data are very influential in reducing their effectiveness. The aim of clustering is to optimize the experimental similarity within the cluster and to maximize the dissimilarity in between clusters. In this study, we use K-means clustering as the basis for the segmentation that will be carried out, and of course, there are additional models that will be used to support the research results. As a result, we have succeeded in dividing the customer into 5 clusters based on the relationship between annual income and their spending score, and it has been concluded that customers who have high-income levels & have a high spending score are also very appropriate targets for implementing market strategies.

Keywords: Segmentation, Strategy, Clustering, K-Means.

1. Introduction

In this era, increasing the level of consumer consumption is very reasonable, this is based on the very fast development of production. This makes each person feel like they have an obligation to spend something to enjoy these developments. At this point, an increase in the number and variance of products is not a bad thing for the market, but an increase in customers can sometimes lead to wasted resources due to a strategy aimed at the wrong customer [1]. At this time a lot of managers and people who work in the marketing field try various things to create the right market strategy. However, we are talking about their customers who are human and change or can change based on various factors. Many applications of certain strategies such as discounts, annual promotions, memberships, etc. may work for a while but after that, it is nothing more than a waste of resources, both energy, and money.

As a manager, it is very important to be able to recognize the patterns or habits of the customers themselves. As a matter of fact, the mall Industry is often involved in a race to increase their customers and therefore make huge profits [2]. There are several factors that explain why the mall rejects its role. First, the level of customer activity is higher, they have less time to shop, and finally, they reduce their shopping frequency. In fact, there are too many of the same malls in a district or city and eventually, customers will go to the shopping center that offers the most products and the best service. This factor encourages mall managers to develop a strategy to differentiate them from competitors [3, 4].

2. Literature Review

research shows that detecting where a customer is going to meet their shopping needs is highly dependent on the service from the provider and the characteristics of the place they are going to. In certain perspectives such

as McGoldrick and Thompson they indicate that the level of price, crowd, convenience, and service are very vital factors. grouping mall classes or categories into functional types, recreation areas, social places & public places is very important. This is used by some mall managers / owners in determining target visitors and the types of services they provide.

Using quite different perspectives, Lehew and Wesley categorize the mall section as multiple neighborhoods and values. Anselmsson said that the most important factor in fulfilling customer satisfaction is the selection of the right atmosphere, equipment, promos, and communicative methods. With this, it is hoped that this type of grouping will provide the best service & experience for customers who come. Previously we mentioned several types of grouping of important factors that malls should pay attention to, but customers must also be able to be grouped. This is necessary because most methods or strategies will fail if the target customer group is wrong. In this research, we will classify mall customers using machine learning methods in order to get a clear visualization of the existing customer groupings.

2.1. Machine Learning

We have often seen the application of machine learning in various fields around us, for example, on Facebook, machine learning helps us to identify ourselves and our friends, or even on YouTube recommending videos based on the things you are interested in. Machine learning itself is generally categorized into 2 types, namely Supervised and Unsupervised learning. Supervised learning is usually used by a data analyst to solve problems such as classification and regression [5, 10], which means that in this case the data there is a target label that you want to predict in the future, for example predicting the value of a student or the number of monthly expenses. On the other hand, in unsupervised learning, its users do not always have a special label or target to predict, for example, clustering, based on its mathematical model, the algorithm in unsupervised learning does not have a target of a variable [6]. For example, we want to classify students based on their learning habits or create clusters based on the number of purchases of a product.

The marketing industry, especially the malls, has tough competition to increase their customers and therefore generate huge profits. To achieve this task, machine learning is already being implemented by many shops and other markets [7]. malls or shopping centers take advantage of the data they get when transacting with their customers and make use of it by developing ML models to target the right ones [8,9]. This not only increases sales & the number of visitors who come but also increases efficiency in doing business.

2.2. Clustering

Clustering is known as a method for identifying common groups in a data set. The entities in each group are comparatively more similar to entities from that group than to other group entities. Since the 70s, cluster-based segmentation has been used very often in various studies involving data, especially in marketing. As stated by [11], that clustering is not a structured method of data analysis, even though it has good flexibility, it really depends on the data or sample used. A statistical approach used by some studies with cluster analysis by [12], it is called the "tandem method" consisting of two processes the first is factor analysis and the second is performing cluster analysis.

This approach has been heavily criticized by several other articles. It is all caused by a key problem in his method, namely preliminary factor analysis can destroy existing cluster structures [13]. As an alternative to the tandem method, hierarchical cluster analysis can be used as an alternative by using binary variables. However, the reliability of this method was highly questioned by many researchers at its time and non-hierarchical methods have been very dominant since the 80s.

3. Method

Table. 1. Dataset Structure

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

First of all the research begins by knowing what kind of data we use, see table 1 for the dataset. The dataset we use is quite simple but very detailed, consisting of customer ID, gender, age, annual income, and purchase score. What is meant by a spending score is the value of how much the customer shopped or spent their money at the mall, the value is on a scale from 1 to 100 (higher means the more is spent). The structure of the dataset has been displayed properly, but what about the contents of the data are there any missing values? We are lucky that there is no missing data in our dataset, see Figure 1 below for the results of the total values.
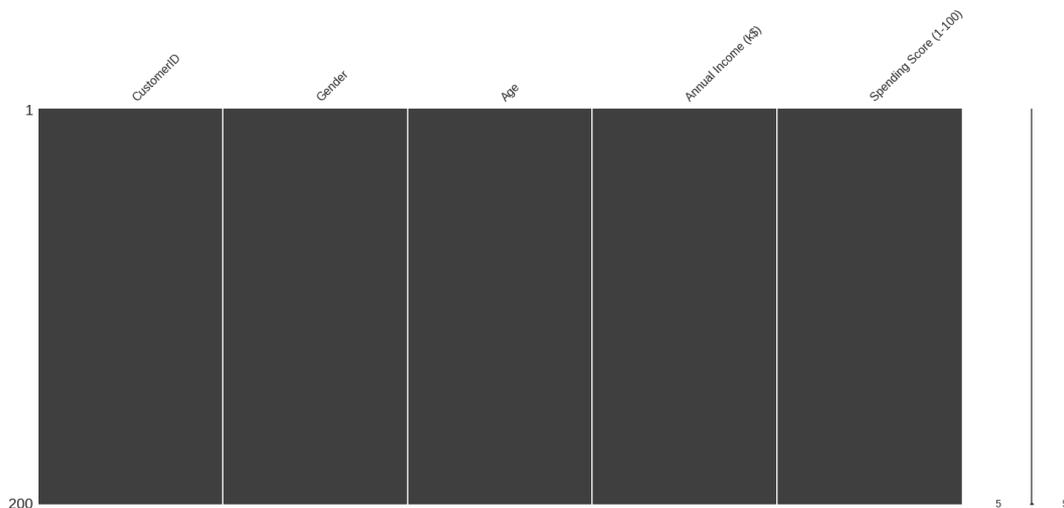

Figure. 1. Dataset Value

After knowing the data we have, we can do the plotting, based on Figure 2 below, we do this by comparing the annual income and spending score, which is of course differentiated by gender. From the results we get we see that there are customer behaviors with annual income and Spending scores, there are 5 type plot shows segments of Customers with the following behaviors:

- Low Income- Low Spending Score
- High Income- Low Spending Score
- Low Income- High Spending Score
- High Income- High Spending Score
- Average Income- Average Spending Score
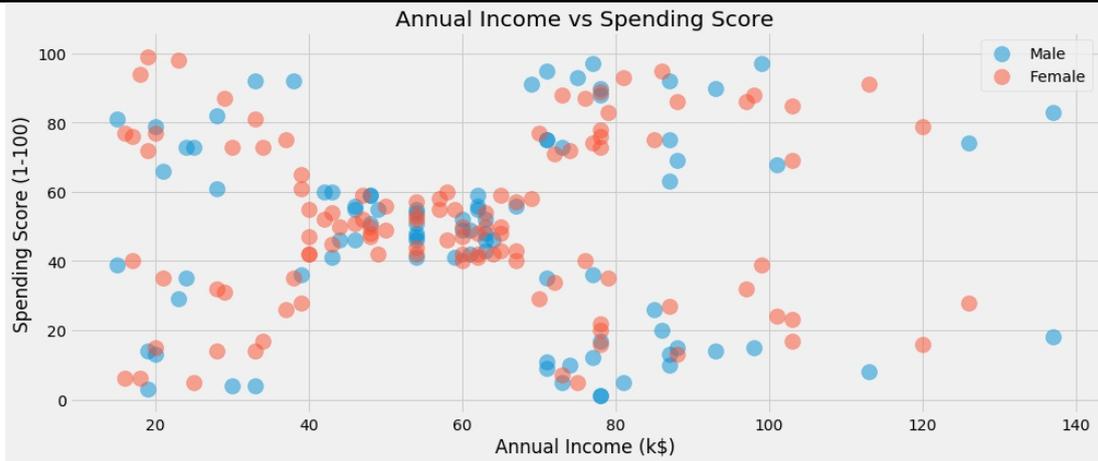
Figure. 2. Annual Income vs Spending Score

Knowing that there are already several groupings, although not in detail we can make a K-means model now. The method we will use is The Elbow method is a really common method as well as the concept would be to perform k-means clustering for a range of k clusters (let's say 1 to 10) and to measure the sum of square distances from each point to its assigned center for each value. The mean squared distance between each instance and its nearest centroid is defined as this [17]. Analytically, the lower inertia of the model is higher, as per the description. We observed that the point after this there is no sudden change in WCSS (Within Cluster Sum of Squares) is found in K=5. And we're going to use K=5 as the right number of clusters that fit the clusters we had before the algorithm was implemented. See figure 3.
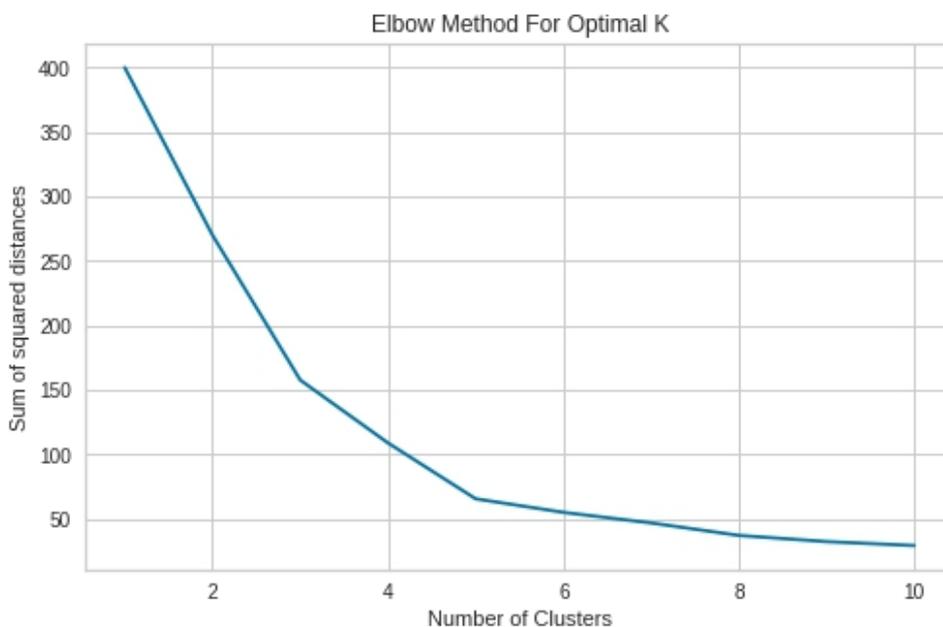

Figure. 3. Elbow method result

By using the above method, we can divide the plot into several clusters and find out which clusters can be prioritized, and give the appropriate label to each cluster. By using the K-means algorithm, we can find out which of the five clusters should be targeted, namely customers with Moderate Income- Moderate Spending Score, High Income- High Spending Score, Low Income- High Spending Score. As we can see in figure 4, the targeted customers have been found.
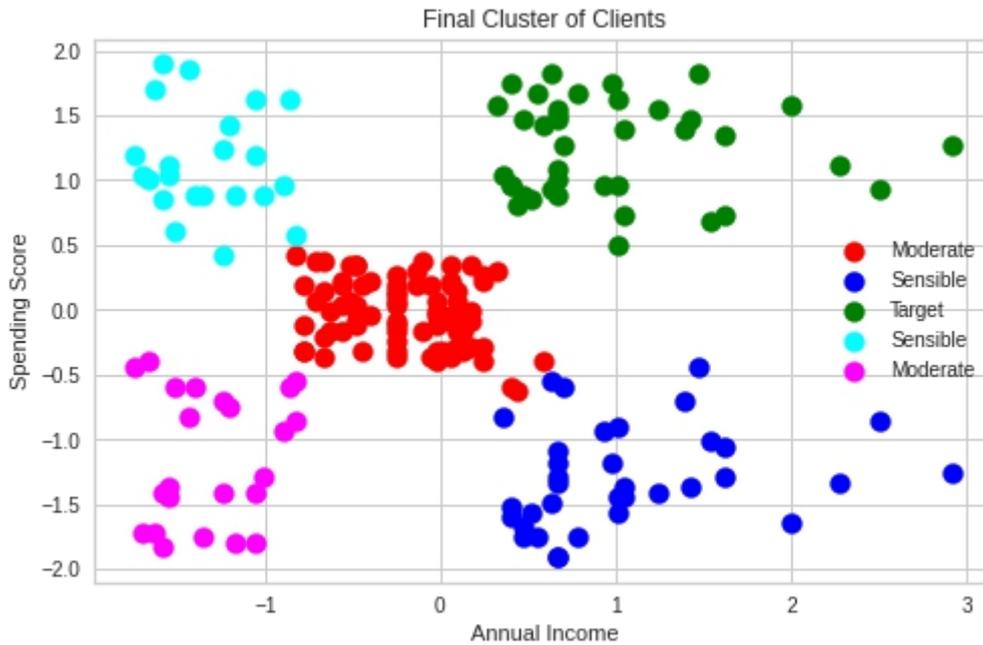
Figure. 4. Final cluster of clients

4. Results and Discussion

We can see that, based on their annual income and spending score, mall customers can be grouped into 5 groups. First, the green group indicates that they are people with high incomes and high spending scores, this is an ideal target for a mall or shopping district because people like this are the biggest and most potent source of profit. In fact, this person may be a regular visitor to a mall and easily be convicted by the mall facilities.

Second, the blue group we can know is a group of people who have a high income but have a low level of spending. This is a very interesting case, given the many potential reasons for such a group to exist. for now, let's assume that they are people who are very active in shopping but are not satisfied with the services or facilities in the mall. such groups are also a good potential target, but we need to be able to identify in advance the reasons for their low levels of spending. the department manager or mall authority can develop or add a facility & offer that can help attract groups of visitors like this to come and have their needs met.

Third, the red group from the data we get identifies them as having average incomes and spending levels. we can assume these are people who don't always buy a product but have a high level of willingness to spend even though sometimes they have little income. This group of people is not a group that has a high potential high income for the mall, and also as a manager as much as possible to avoid targeting this group of people in a market strategy. however, they can still be considered through other data analysis techniques that might increase their level of spending.

fourth, the cyan-colored group, as we can see, this group contains people who have low income but have high spending scores, people like this have pleasure or hobby in spending something even though they have a low income. This is also possible if they are people who feel comfortable or satisfied with the services provided by the mall so that they feel compelled to spend something because the service makes them satisfied.

Fifth, the yellow group classifies people who have low annual incomes and low spending scores. and it is quite reasonable also that they have low income so they will spend less on something, even what they do maybe a wise and good choice based on their condition. a mall manager should target the people in this cluster at the lowest priority.

Based on their Annual Income and Spending Score, we know the behavior of customers by looking at the results. Many marketing tactics for customers can be adapted to this cluster study. Our target customers are high income and high spending score customers, and we will still like to keep them as they offer the most profit margin. Customers with a wide range of products will be attracted to their lifestyle requirements for high income and lower spending score and that could attract them to the Mall Supermarket. Less Income Less Spending Score can be sent additional promotions and they will be drawn to spending by continuously giving them offers and discounts. A cluster analysis can also be performed on what kind of products consumers choose to consume and can find other marketing campaigns accordingly.

5. Conclusion

This research proves that it is possible to do segmentation on customers in malls. even the application of machine learning like this is very profitable in the industry, a manager can pay full attention to handling each cluster that has been identified by meeting their every need. To meet the needs of customers, mall managers must be able to understand what is needed and be in the minds of customers, study their shopping habits and maintain regular interactions with customers that can make them feel comfortable.

This research proves that it is possible to implement machine learning in the industrial segmentation of this shopping district. But assuming machine learning can perform clustering with fairly accurate accuracy may still be extremely difficult to fully implement permanently. because even though the data we get comes from customers and is structured, we are talking about humans, they can learn, and of course, changing a habit or changing their spending patterns is something they might do. Assuming that implementing clustering like this can give wrong results, it is safer to still let a manager make decisions in determining a target or strategy. however, this does not close the answer that its application failed as the fact that the results we get in this study can be arguably appropriate for use. The application of machine learning in this study may open up the potential for other applications in the same industry

References

[1] C. Calvo-Porral and J. P. Lévy-Mangin, "Profiling shopping mall customers during hard times," *J. Retail. Consum. Serv.*, vol. 48, no. November 2018, pp. 238–246, 2019, doi: 10.1016/j.jretconser.2019.02.023.

[2] A. G. Parsons, "Assessing the effectiveness of shopping mall promotions: Customer analysis," *Int. J. Retail Distrib. Manag.*, vol. 31, no. 2, pp. 74–79, 2003, doi: 10.1108/09590550310461976.

[3] L. Lucia-Palacios, R. Pérez-López, and Y. Polo-Redondo, "Does stress matter in mall experience and customer satisfaction?," *J. Serv. Mark.*, vol. 34, no. 2, pp. 177–191, 2020, doi: 10.1108/JSM-03-2019-0134.

[4] M. F. Diallo, F. Diop-Sall, S. Djelassi, and D. Godefroit-Winkel, "How Shopping Mall Service Quality Affects Customer Loyalty Across Developing Countries: The Moderation of the Cultural Context," *J. Int. Mark.*, vol. 26, no. 4, pp. 69–84, 2018, doi: 10.1177/1069031X18807473.

[5] L. Kouhalvandi, O. Ceylan, and S. Ozoguz, "Automated Deep Neural Learning-Based Optimization for High Performance High Power Amplifier Designs," *IEEE Trans. Circuits Syst. I Regul. Pap.*, pp. 1–14, 2020, doi: 10.1109/tcsi.2020.3008947.

[6] S. Hidayat, M. Matsuoka, S. Baja, and D. A. Rampisela, "Object-based image analysis for sago palm classification: The most important features from high-resolution satellite imagery," *Remote Sens.*, vol. 10, no. 8, 2018, doi: 10.3390/RS10081319.

[7] L. L. Rego, N. A. Morgan, and C. Fornell, "Reexamining the market share-customer satisfaction relationship," *J. Mark.*, vol. 77, no. 5, pp. 1–20, 2013, doi: 10.1509/jm.09.0363.

[8] X. Luo and C. B. Bhattacharya, "Social Responsibility , Corporate Customer and Market Satisfaction , Value," *Am. Mark. Assoc.*, vol. 70, no. 4, pp. 1–18, 2006.

[9] A. Doulamis and N. Doulamis, "Customer Experience Survey," pp. 3–6, 2016, doi: 10.3390/technologies8040076.

[10] Akmal, "Predicting Dropout on E-learning Using Machine Learning," *J. Appl. Data Sci.*, vol. 1, no. 1, pp. 29–34, 2020.

[11] J. P. Ruiz, J. C. Chebat, and P. Hansen, "Another trip to the mall: A segmentation study of customers based on their activities," *J. Retail. Consum. Serv.*, vol. 11, no. 6, pp. 333–350, 2004, doi: 10.1016/j.jretconser.2003.12.002.

[12] O. Dogan, C. Fernandez-Llatas, and B. Oztaysi, *Process mining application for analysis of customer's different visits in a shopping mall*, vol. 1029. Springer International Publishing, 2020.

[13] K. jae Kim and H. Ahn, "A recommender system using GA K-means clustering in an online shopping market," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1200–1209, 2008, doi: 10.1016/j.eswa.2006.12.025.

[14] LeHew, M. L., & Wesley, S. C. (2007). Tourist shoppers' satisfaction with regional shopping mall experiences. International Journal of Culture, Tourism and Hospitality Research, 1(1), 82-96.

[15] Anselmsson, J. (2006). Sources of customer satisfaction with shopping malls: a comparative study of different customer segments. International Review of Retail, Distribution and Consumer Research, 16(1), 115-138.

[16] McGoldrick, P. J., and Thompson, M. G. (1992). The role of image in the attraction of the out-of-town centre. International Review of Retail, Distribution and Consumer Research, 2(1), 81-98.

[17] M. Imron, U. Hasanah, and B. Humaidi, "Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping," IJIIS Int. J. Informatics Inf. Syst., vol. 3, no. 1, pp. 12–22, 2020, doi: 10.47738/ijiis.v3i1.3.

[18] Akmal, "Predicting Dropout on E-learning Using Machine Learning," J. Appl. Data Sci., vol. 1, no. 1, pp. 29–34, 2020, [Online]. Available: http://bright-journal.org/Journal/index.php/JADS/article/view/6.