ANALYSIS OF MARKETING SEGMENTATION AND ITS IMPLEMENTATION ON 7PS ERHA'S TREATMENT ULTIMATE ACNE CURE USING K-MEANS CLUSTERING

Alisa Agustine, Hardiyanti, Lisye Ira Anne, Jerry Heikal

Magister of Management Program, Universitas Bakrie, Indonesia
Email: agustinealisa@gmail.com, hardiyantis123@gmail.com, lisyeiraanne@gmail.com, jerry.heikal@bakrie.ac.id

Abstract

K-Means is a non-hierarchical data grouping method that separates existing data into two or more groups. This method separates existing data into groups so that data with the same character is included in the same group and data with different characters is grouped into other groups. This study aims to produce an analysis that can classify Erha Clinic product/treatment data for March 2022 period using software SPSS 25 IBM to make marketing strategies more targeted. This study divided Erha Clinic product/treatment data with attributes of age group, type of plan, gender, education and total purchase into three clusters (Cluster Metrosexual, Cluster Millenials and Cluster Generation Z). The clustering process is 10 iterations with a minimum distance between clusters of 8,746. The significance value indicates that there is a significant difference between clusters 1, cluster 2 and cluster 3 related to Gender as one of the attributes in the study. The results of clusters show that the marketing target chosen by Erha Clinic is in cluster 3 (Gen-Z Persona) due to acne problems are mostly experienced by young people and although the transaction price is cheap, in the cluster 3 has the most purchased acne products compared to the purchase of the Advance plan and Basic plan bundling in cluster 1 and cluster 2.

Keywords: K-Means, Cluster, Iteration, Significant.

Introduction

PT. Arya Noble is a strategic holding company and investment based in Indonesia. Arya Noble has several business units engaged in pharmaceutical companies, such as Genero (a manufacturer of skincare & medicine), Dermies (Beauty clinic), Skinproof (research agency), and several other companies, Erha Clinic. Erha is one of the subsidiaries

of the company that focuses on health in Indonesia and specializes in dermatology (skin and hair) beauty in Indonesia.

Erha Clinic is engaged in personal care which is closely related to dermatology, from oil gland problems, facial skin problems, scalp to hair health. As quoted from Erha's official website, Erha.co.id, Erha believes that skin is an important part of appearance and can affect one's self-confidence. As the best skin clinic, Erha believes that everyone deserves healthy skin. That is why Erha was founded with the aim of being a solution for various skin disorders, which will be supported by the best specialist doctors, quality products and treatments.

Erha Clinic Indonesia itself is divided into 2 major categories, namely the skincare category, which are skin and hair health products that can be purchased by consumers without having to use a doctor's prescription. While the other one is the Clinical Program category, which is a personalized concept category where consumers will get a series of products and treatments that are tailored to skin and hair conditions and problems (Puebla-Barragan & Reid, 2021).

The Clinical Program category itself is divided into 6 brands including: Ultimate Acne Cure, Ultimate Anti-Aging, Ultimate Brightening, Ultimate Hair Care, Ultimate Make Over dan Ultimate Atopy Cure. Of the 6 brands in the clinical program category, Ultimate Acne Cure is the largest contributor to sales contribution to total revenue. Thus, in order for the business to run sustainably and continue to grow, it is necessary to carry out marketing in order to continue to attract the attention of target customers.

Through the existing customer data of Ultimate Acne Cure patients, it can be used as material for analysis so that it can be taken into consideration whether the marketing that has been done is correct, or needs to be improved. One way of analyzing the data is by grouping K-Means Clustering data. The K-Means algorithm is a non-hierarchical data clustering method that tries to divide data into one or more clusters. Data that has the same characteristics are in one cluster and data that has different characteristics are grouped in another group.

Based on the description above, this study will discuss the use of the K-Means algorithm to determine the product/treatment purchasing cluster presented by Erha Clinic with age group variables, type of plan, gender, education and total purchases, so that the marketing department will know what marketing mix strategy is right for selling Erha Clinic products / treatments.

Based on the identification of the problems that have been described, the problem in this study is formulated as follows: How to apply the K-Means algorithm to determine the cluster of product / treatment purchases presented by Erha Clinic with variables of age group, type of plan, gender, education and total purchases during March 2022 period.

The limitations of the discussion in this study are; The data processed is Erha Clinic transaction data for the period March 2022. The method used is the K-Means method. The

data was processed using IBM's SPSS 25.

Literature Review
A. Clustering

Clustering is a method for finding and grouping data that have similar characteristics (similarity) between one data and another. Clustering is a data mining method that is unsupervised. In data mining there are two types of clustering methods used in data grouping, namely hierarchical clustering and non-hierarchical clustering (Santosa, 2007).

Hierarchical clustering is a data grouping method that starts by grouping two or more objects that have the closest similarity (C. Wu et al., 2021). Then the process is passed to another object which has a second immediacy. And so on so that the cluster will form a kind of tree where there is a clear hierarchy (level) between objects, from the most similar to the least similar. Logically, all objects in the end will just form a cluster. The non-hierarchical clustering method begins by determining in advance the desired number of clusters (two clusters, three clusters, or so on). After the number of clusters is known, then the cluster process is carried out without following the hierarchical process. This method is commonly called K-Means Clustering (Santoso, 2010).

B. K-Means Method

The K-Means algorithm is an iterative clustering algorithm that partitions the set into a number of K clusters that have been set at the beginning. The K-Means algorithm is simple to implement and run, relatively fast, easy to adapt, and has been widely used in practice historically (X. Wu & Kumar, 2009).

According to Santosa (2007), the steps for clustering with the K-Means method are as follows:

1. Data Standardization

If the number of variables is far enough from one variable to another which can complicate the grouping process which makes the data invalid, the parameter does not dominate in calculating the distance between data and creates duplicated data. If it has significantly different units, standardize the data using the Z-Score formula so as to produce a balance of comparison values between the data before and after the process. Standardization of data is done by using the following formula:

$$z_i = \frac{x_i - \bar{x}}{S}$$

Information:

z_i = Value Z- Score to-i

x_i = Value Datum to-i

x = Average Value

S = Value Standart Deviation

2. Determine The Number of Clusters-K.

In this study, the number of clusters was divided into three clusters, namely cluster 1, cluster 2, and cluster 3, namely determining the age group, type of transaction plan gender, education level, and the total amount of transaction payments from the highest/highest to the least/lowest.

3. Determine the center point or centroid with the help of IBM's SPSS 25 application.

4. Calculates the distance to the center of the group. The distance between the data and the centroid is done using the theory of Euclidean distance with the formula used is the following formula:

$$d_{ij} = \sqrt{(x_{1j} - y_{1k})^2 + (x_{2j} - y_{2k})^2 + \cdots + (x_{ij} - y_{jk})^2}$$

$$d_{ij} = \sqrt{\sum_{i=1}^{p} (x_{ik} - x_{jk})^2}$$

Information:

d_{ij} = distance between x_i and x_j

p = variable cluster distance

x_{ik} = shows the data value from i point to k dimension

x_{jk} = shows the data value of the initial center of the cluster from the j point to k dimension

5. Group each data to the closest distance to the center.

6. The reallocation of data into each group into K-Means is based on the comparison of the distance between the data and the centroid of each existing group. This allocation can be done with the following formula:

$$a_{ij} = \begin{cases} 1 \; ; d = \min \{d(x_{i,} c_1)\} \\ 0 \; ; \text{Lainnya} \end{cases}$$

a_{ij} is the membership value point x_i to centroid C_1, d is the shortest distance from data x_i to group k after comparison, and C_1 is centroid ke-1

7. Determine the position of the new cluster center.

8. The new cluster center or C_{kj} by calculating the average value of the data in the same cluster with the following formula:

$$C_{kj} = \left(\frac{1}{n_k}\right) \sum_{l=1}^{p} x_{jl}$$

C_{kj} = New cluster center to-k on variable to-j

nk = The number of object members in the cluster to-k

x_{jl} = Data in cluster to-j on variable to-l

9. If the cluster center does not change again, the cluster process is complete, or return

to step 3 if there is still data moving clusters.

C. Marketing Mix

According to Kotler (2009) that Marketing Mix is a set of marketing tools that companies use to continuously achieve their marketing goals in the target market. On the other hand, there are adjustments to the marketing mix, where the producer adjusts the elements of the marketing mix for each target market. The variables in the marketing mix can be used effectively if they are arranged according to the circumstances and situations that are being experienced in a company.

From the above definition it can be concluded that the notion of the marketing mix is the factors that are controlled and can be used by marketing managers to influence consumer purchasing decisions. These factors include Product, Price, Place, Promotion, People, Process and Physical Evidence.

1. Product

According to Kotler & Armstrong (2001) "Product as anything that can be offered to a market for attention, acquisition, use, or consumption and that might satisfy a want or need". A product is anything that a producer can offer to be noticed, requested, sought, purchased, used, or consumed by the market as a fulfillment of the needs or desires of the relevant market, either in the form of goods or services. Products can be measured including through Kotler, (2005): (a) Product Variation. (b) Product quality. (c) Product display

2. Promotion

Promotion is a company's effort to influence potential buyers through the use of all elements or the marketing mix (7P) (Rachmawati et al., 2021). Promotional media that can be used in this business include advertising, sales promotion, publicity and public relations, and direct marketing (Camilleri & Camilleri, 2018). The determination of the promotional media to be used is based on the type and form of the product itself. Promotion can be broadly measured through Tjiptono (1995): a) Ad attractiveness rate. b) Competitor publicity.

3. Price

Price has a major role in the decision-making process of consumers (Tjiptono, 1995). The price depends solely on the company's policy, but of course taking various things into account. The price is said to be expensive, cheap, or mediocre for each individual, it does not have to be the same, because it depends on the individual who is motivated by the environment and individual conditions. According to Chandra (2002) prices can also be measured including through: a) Competitive product prices. b) Discount (discounted price). c) Payment system variations.

4. Place

According to Sutojo (2009) distribution is an effort so that a product can be available in places that make it easier for consumers to buy it whenever consumers need it. Site selection requires careful consideration of several factors, including: a) Access, for example a road that makes it easier for consumers to reach the place. b) Visibility, for example a location that can be seen clearly from the side of the road. c) Parking lots, have their own parking space or space or use public parking lots. d) Expansion, there is sufficient space for business expansion in the future. e) Government regulations, such as business licenses. f) Competition, namely the consideration of competitors' locations.

5. People

According to Ratih (2015), people are: "all actors who play a role in the presentation of services or products so that they can influence purchases". The elements of people are company employees, consumers and other consumers in the service environment. According to (Hurriyati, 2015) this people element has 2 aspects, namely:

a. Service People

For service organizations, service people usually hold dual positions, namely providing services and selling those services. Through good, fast, friendly, thorough and accurate service, it can create customer satisfaction and loyalty to the company which will ultimately improve the company's good name.

b. Customer

Another influencing factor is the relationship that exists between the customers.

6. Process

According to Philip Kotler (2006), the process here includes how the company serves the demands of each customer. Starting from the consumer ordering (order) until they finally get what they want. Certain companies usually have a unique or special way of serving their customers. What is meant by the process in marketing is the whole system that takes place in the implementation and determines the quality of the smooth operation of services that can provide satisfaction to its users.

7. Physical Evidence

According to Ryu (2011) physical facilities are very important for restaurants because they support the atmosphere in the restaurant which can affect the enjoyment obtained by consumers. Physical facility indicators are classified into six variables, namely (Liu et al., 2021): a) Color. b) Layout. c) Lighting. d) Facilitating goods. e) Furnishing

Research Methods

The research stages used in the Implementation of Erha Ultimate Acne Cure 7Ps Marketing Strategy Analysis Using K-Means Clustering, are shown in Figure 1:
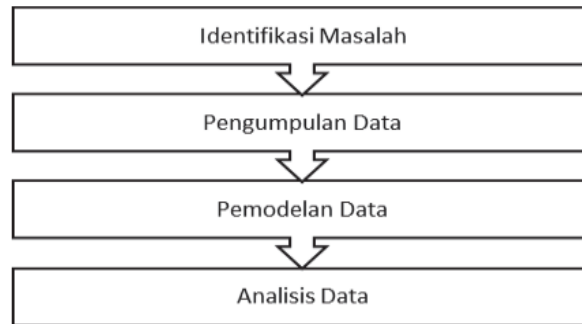
Figure 1 Research Stages

Phase 1 is problem identification, based on the results of transaction survey data for the March 2022 period, customers from the Erha Ultimate Acne Cure program tend to vary. Proper management of survey results is expected to produce sales targets for other products in the Ultimate Acne Cure program in order to find out the right marketing target segmentation. Therefore, this research was made to provide information that will later support the marketing strategy, so that the promotional activities of the Erha program become more efficient based on existing data.

Stage 2 is data collection, the data needed in this study was obtained through the results of a marketing team survey from Erha with 6750 product/treatment data, male and female gender, types of plans in the form of Advance, Basic and unit product, Education starting from Elementary School to Strata II, the age group starts from 0 - 71+ years and the distribution of the number of transactions starts from high to low.

Stage 3 is data modeling, the previous data is data that we cannot process because it is still in the form of characters, K-Means Clustering is an algorithm that can only work when the processed data is data in the form of numbers or integers. So, the above data must be initialized so that it can be analyzed using the K-Means Clustering algorithm. Before the initialization step, the K-Means Clustering analysis process can be seen in Figure 2.
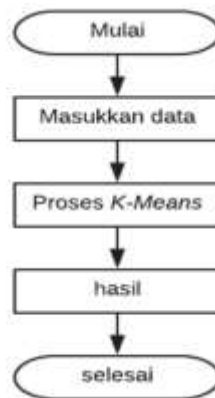
Figure 2 Flowchart K-Means Clustering

Figure 2 illustrates the flow of the system running in a system that was built to display the results of data analysis using the K-Means Clustering Algorithm for the Erha Ultimate Acne Cure Program marketing strategy. The data used is secondary data from the results of a direct customer survey of Erha Ultimate Acne Cure users in the March 2022 period. Data processing is assisted by the IBM SPSS 25 (Statistical Package for the Social Science) application. Starting from importing data in xls or xlsx format, after that the input data must be processed first by initializing the data based on frequency (large to small). The initialized data is processed using the K-Means Clustering Algorithm so that the final result of the data analysis is in the form of a report showing the data grouping formed.

Results and Discussion

This research used software SPSS 25 IBM in which it has the results as below:

Table 1
Descriptive Statistics

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Age (group)2	6750	1	10	4.00	1.847
Jenis Plan	6750	1	3	2.21	.706
Gender	6750	1	2	1.15	.362
Pendidikan	6750	1	9	4.81	2.138
Sum of Revenue	6750	201000	17992800	552962.82	710613.445
Valid N (listwise)	6750				

Based on Table 1 Descriptive Statistics, it shows Minimum, Maximum, Mean and Standard Deviation with complete data of 6750 product / treatment data based on age group, type of plan, gender, education and sum of revenue.

Tabel 2
Initial Cluster

Alisa Agustine, Hardiyanti, Lisye Ira Anne, Jerry Heikal

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore: Age (group)2	1.62400	1.08248	-1.62512
Zscore: Jenis Plan	1.11295	1.11295	-1.71979
Zscore(Gender)	-.42771	-.42771	2.33767
Zscore(Pendidikan)	1.02472	1.02472	-1.78151
Zscore: Sum of Revenue	24.54195	15.81315	-.46152

Based on table 2. Initial Cluster Centers, in this table it shows the initial step of the formation of the three clusters. After that, K-Means Cluster method will execute the test and iteration and iteration for data relocation, therefore there is no object that will move from one cluster to another cluster because later on, it will execute clustering process after the iteration which will be final cluster results, so, this output will not be analyzed.

Tabel 3
Iteration History

Iteration History[a]

Iteration	Change in Cluster Centers		
	1	2	3
1	.000	5.217	3.797
2	.000	2.499	.014
3	.000	1.785	.017
4	4.373	1.508	.030
5	1.983	1.318	.058
6	4.113	.720	.053
7	.603	.268	.022
8	.598	.127	.010
9	1.493	.079	.003
10	.000	.035	.003

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is .021. The current iteration is 10. The minimum distance between initial centers is 8.746.

From the output result of SPSS 25 IBM in the Table 3 Iteration History, it is known that the iteration process is performed ten times. This process is executed to obtain the right cluster in terms of clustering product / treatment. The minimum distance between initial clustes center is 8.746.

Then, for the next step, the result of K-Means is final cluster centers. There are three

clusters in table 4 which divide consumers' transaction data regarding product / treatment in Erha Clinic based on age group, type of plan, gender, education and sum of revenue. Output of Final Cluster Centers is still related to the standardization process of the previous data (Zscore).

Tabel 4
Final Cluster Centers

Final Cluster Centers

	Cluster		
	1	2	3
Zscore: Age (group)2	.65700	.27342	-.02359
Zscore: Jenis Plan	.30360	.38510	-.03183
Zscore(Gender)	.16487	.05513	-.00483
Zscore(Pendidikan)	.52361	.20810	-.01801
Zscore: Sum of Revenue	11.74669	2.49522	-.22818

Based on table final cluster centers for product / treatment in Erha Clinic, it is obtained the results as below:

1. In the cluster 1 consists of an older age group, transaction type tends to basic plan with mostly male gender, high level of education and expensive transaction price. The persona in cluster 1 is called Persona Metrosexual.
2. In the cluster 2 consists of middle age group, transaction type tends to advance plan with mostly female gender, middle level of education and middle transaction price. The persona in cluster 2 is called Persona Millennials.
3. In the cluster 3 consists of young age group, transaction type tends to product plan where they only buy product in unit and not in bundling type with mostly female gender, low level of education and cheap transaction price. The personal in cluster 3 is called Personal Generation Z

Tabel 5
Tabel ANOVA

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore: Age (group)2	23.595	2	.993	6747	23.754	.000
Zscore: Jenis Plan	41.173	2	.988	6747	41.669	.000
Zscore(Gender)	1.029	2	1.000	6747	1.029	.357
Zscore(Pendidikan)	13.842	2	.996	6747	13.895	.000
Zscore: Sum of Revenue	2697.127	2	.201	6747	13432.410	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

After forming 3 clusters, the next step is to see whether the variables that have formed clusters have differences in each cluster. In this case, it can be seen from F and the probability value (sig) of each variable. This is done by looking at the Anova output. The interpretation of the F number is that the greater the F number of a variable and the significance number is 0.05, the greater the difference between these variables in the five variables. For example, the largest F number (13432.41) is in the Z Sum of Revenue, with column numbers Sig. 0.000 which means the significance is real. This means that the Sum of Revenue factor greatly distinguishes the characteristics of the three clusters.

Or it can be said that the Sum of Revenue in the three clusters is very different between cluster 1 and other clusters. In the variable Z, Age of Group has an F number of 23,754 and Sig 0.00, which means the significance is also real. In variable Z, Plan Type has an F number of 41,669 with Sig 0.00 which means the significance is real. For variable Z Education has an F number of 13,895 with Sig 0.00 which means the significance is real. If you pay attention to the difference with the Z Gender variable, which has an F number of 1.029 with a sig of 0.357, it states that the significance is above 0.05 (0.357 > 0.05). Then the variable Z Gender in cluster 1, cluster 2 and cluster 3 has a difference

.

Tabel 6
Number of Cases in each Cluster

Number of Cases in each Cluster		
Cluster	1	14.000
	2	504.000
	3	6232.000
Valid		6750.000
Missing		.000

In Table 6, it can be seen that the most product/treatment purchase data is in cluster 3, which is 6751 product/treatment data. The purchase data for intermediate products/treatments is in cluster 2, which is 504 product/treatment data. Meanwhile, data on

product/treatment purchases are at least in cluster 1, which is 14 product/treatment data. Because there are no missing variables, thus all the 6750 complete product/treatment data are recorded in the 3 clusters with the composition as above because cluster 3 is the largest cluster. So from the number of cases in each cluster, it can be concluded that the marketing target chosen by Erha Clinic is in cluster 3 (Persona Gen-Z) with the reason that acne problems are mostly experienced by young people and even though the transaction price is cheap, buying acne products the most compared to the purchase of Advance plan and Basic plan bundling in cluster 1 and cluster 2.

Conclusions

Based on the analysis results of the data grouping of Erha Clinic product/treatment data, the following conclusions are: (1) the number of clusters is three clusters based on the number of purchases of Erha Clinic products/treatments with variables of age group, type of plan, gender, education, and sum of revenue. (2) In cluster 1 there is an older age group, transaction types tend to be basic plan types with mostly male gender, high education level and expensive transaction prices. (3) In cluster 2 there is a middle age group, the type of plan transactions tend to be the type of advance plans with mostly female gender, education level and medium transaction prices. (4) In cluster 3, there are young age groups, plan product type transactions where only unit purchases are made and not packages with mostly female gender, low level of education and low transaction prices. (5) So that the marketing target chosen by Erha Clinic is in cluster 3 (Persona Gen-Z) with the reason that acne problems are mostly experienced by young people and even though the transaction price is cheap, the purchase of acne products is the most compared to the purchase of the Advance plan and Basic bundling. plan in cluster 1 and cluster 2.

Alisa Agustine, Hardiyanti, Lisye Ira Anne, Jerry Heikal

BIBLIOGRAPHY

Camilleri, M. A., & Camilleri, M. A. (2018). *Integrated marketing communications*. Springer.

Chandra, G. (2002). Strategi & Program Pemasaran. *Yogyakarta: Andi Offset*.

Hurriyati, R. (2015). *Bauran pemasaran dan loyalitas konsumen: Fokus pada konsumen kartu kredit perbankan*.

Kotler, P. (2005). Manajamen pemasaran, jilid 1 dan 2. *Jakarta: PT. Indeks Kelompok Gramedia*.

Kotler, P., & Amstrong, G. (2006). Dasar-Dasar Pemasaran Edisi XII. *PT. Indeks, Jakarta*.

Kotler, P., Armstrong, G., Saunders, J., & Wong, V. (2001). Principal of marketing. *City: Harlow Pearson Education Limited*.

Kotler, P., & Keller, K. (2009). Manajemen Pemasaran. 13th. *Edisi Bahasa Indonesia. Jakarta: PT. Indeks*.

Liu, J., Kamarudin, K. M., Liu, Y., & Zou, J. (2021). Developing pandemic prevention and control by ANP-QFD approach: A case study on urban furniture design in China communities. *International Journal of Environmental Research and Public Health*, *18*(5), 2653.

Puebla-Barragan, S., & Reid, G. (2021). Probiotics in cosmetic and personal care products: Trends and challenges. *Molecules*, *26*(5), 1249.

Rachmawati, E. R., Syah, T. Y. R., Indradewa, R., & Fajarwati, D. (2021). Influence of Marketing Mix Strategy on Business Arena Corner. *International Journal of Research and Review*, *8*(8), 76–86.

Ryu, K., & Han, H. (2011). New or repeat customers: How does physical environment influence their restaurant experience? *International Journal of Hospitality Management*, *30*(3), 599–611.

Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta: Graha Ilmu*, *978*(979), 756.

Santoso, S. (2010). *Statistik multivariat*. Elex Media Komputindo.

Sutojo, S. (2009). Manajemen pemasaran. *Jakarta: PT. Damar Mulia Pustaka*, *1*(2).
Tjiptono, F. (1995). *Strategi pemasaran*. Andi Offset.

Wu, C., Peng, Q., Lee, J., Leibnitz, K., & Xia, Y. (2021). Effective hierarchical clustering based on structural similarities in nearest neighbor graphs. *Knowledge-Based Systems*, *228*, 107295.

Wu, X., & Kumar, V. (2009). *The top ten algorithms in data mining*. CRC press.