

Date: Monday, August 24, 2020 Statistics: 197 words Plagiarized / 2453 Total words Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

Cosmetics Customer Segmentation and Pro?le in Indonesia Using Clustering and Classi?cation Algorithm Sari Hartini 1 \* , Windu Gata 2 , Sigit Kurniawan 3 , Hendra Setiawan 4 and Kadinar Novel 5 1Technology Informatics, STMIK Nusa Mandiri, Jakarta 2Computer Science, STMIK Nusa Mandiri, Jakarta 3Technology Informatics, STMIK Bani Saleh, Bekasi 4System Informatics, STMIK Bani Saleh, Bekasi 5Technology Informatics, Universitas Bina Sarana Informatika, Jakarta E-mail: sari.shi@nusamandiri.ac.id Abstract.

The cosmetics business competition in Indonesia is currently increasing so rapidly, cosmetics customers have spread to various brands, and according to taste. The customer for the company is an asset that is very important for business continuity, so that good customer management can increase company revenue. However, it is not easy to manage customers if they cannot read the characteristics of customers, to carry out appropriate business strategies.

So that requires a customer analysis method that can provide recommendations for the company. RFM is one of the most widely used analytical methods for analyzing customers through segmentation and pro?ling of customers. In addition to segmenting, the customer pro?le is also a very important factor in analyzing customers, ALC is a form of a customer pro?le that can be used.

RFM + ALC method is not easy to do with very large customer history data, so data mining is needed to help conduct the RFM + ALC analysis. Data mining methods using the clustering function with K-Means and the use of the Elbow method to get the most optimal amount of K in the clustering process can be a model used to segment with RFM, as well as the Na<sup>"</sup>ive Bayes and Decision Tree classi?cation methods to determine ALC pro?le factors the most in?uential customer.

The results of clustering modeling carried out produce two dominant customer segments. While the Na<sup>"</sup>ive Bayes classi?cation model of the ALC factor can provide recommendations for the most in?uential customer pro?les, with the highest level of accuracy with an accuracy value of 65.87% when compared to the Decision Tree. 1.

Intro duction Business competition is now a concern for every company organization, marketing strategies, good marketing policies, getting new customers, and retaining old customers are very important things. The organization will use strategies that provide the best satisfaction for customers and which will return investment in the form of pro?ts [1].

The customer is an important asset for the sustainability of the company's organization [2], good customer management will a?ect revenue for the company. Customers are the determinants of success or failure of a business organization and without customers, it is impossible to form a business. Doing maintenance for customers is an important part, in addition to getting new customers.

Not a few companies that lose a lot of customers because they can not maintain their customers by providing good service. Incorrect in determining strategies related to marketing to customers is one of the causes of loss of customers for the company. In the world of marketing, there is a strategy that is widely used to build relationships between companies and customers, commonly known as Customer Relationship Management (CRM) [3], [4]. Enhancing CRM strategies in companies requires support for data analysis that can mine patterns and trends from their customers [5].

One way that can be used to support the company's strategic needs to improve customer development is to do pro?ling or drawing customer data and customer segmentation [6]. The analytical method that can be used for these needs and has been widely used is RFM (Recency, Frequency, Monetary), a method of analyzing customer behavior based on the history of their transaction updates, a number of transactions and transaction value carried out with the aim of providing recommendations for CRM strategies in the company [6].

However, it is not easy for companies to do segmentation and pro?ling based on customer behavior from large databases for various needs related to market promotion, or achieving sales targets using the RFM analysis method. This research takes an object in a cosmetics retail company in Indonesia, namely Martha Tilaar Shop with RFM analysis method combined with Age, Location, and Cellular Operator (Celular), or can also be called RFM + ALC.

The object was chosen because, in the cosmetics company, customer data analysis has not yet fully used an e?ective data analysis model in determining business strategies for customers. The use of data mining as a method that is currently widely used to dig up the information contained in large data is very helpful, one of which is the problem of customer segmentation and pro?ling. The need to obtain an e?ective analysis model for segmenting and pro?ling customers is very important to be applied.

Clustering and classi?cation techniques can be a way of segmenting and pro?ling these customers. 2. RFM and Segmentation Customer 2.1. RFM (Recency, Frequency, Monetary) Analysis RFM analysis according to Margaret Rouse in the online article de?nes RFM analysis is a marketing technique used to quantitatively determine which customers are best by checking how recently customers buy (Recency), how often they buy (Frequency), and how much customers spend (Monetary).

RFM analysis is based on the marketing axiom that "80% of your business comes from 20% of your customers [7]. RFM analysis is a proven marketing model for customer segmentation based on behavior. It groups customers according to their transaction history - how new, how often, and how much they buy.

RFM helps divide customers into various categories or groups to identify customers who are more likely to respond to promotions and also for personalized services in the future [8]. RFM is a method used to analyze customer value. This is commonly used in database marketing and direct marketing and has become a particular concern in the retail and professional services industries [9].

RFM consists of three dimensions, including: • Recency - How long did the customer last buy to date. • Frequency - How many customers buy. • Monetary Value - How much total money the customer spends. 2.2. Customer Segmentation & Pro?ling According to Philip Kotler and Gary Armstrong (2004), the notion of market segmentation (customers) is the division of a market into several distinct buyer groups.

The purpose of market grouping is to divide markets that are di?erent (heterogeneous) into homogeneous market groups, where each group can be targeted to market a product according to the needs, desires, or characteristics of buyers in the market [10]. Customer segmentation according to Santi Meitasari. ST, M.Entr quoted from the online page of a website, is an attempt to group customers into di?erent groups based on needs, behavior, and other traits.

Customer segments can also be de?ned through demographics such as age, ethnicity,

profession, gender, etc. or their psychographic factors such as shopping behavior, interests, and motivation [11]. According to Swastha and Handoko (2000), the notion of market segmentation is the activity of dividing a market that has heterogeneous nature into a one-unit market that is homogeneous.

According to Marina Silalahi in an online article mentions four types of e?ective customer segmentation, including [12]: (i) Demographic segmentation is a grouping of customers using demographic data or their population using measurement tools such as age, gender, education, employment, income, marital status. (ii) Geographical segmentation, which involves grouping customers according to country, state, region, climate or market size.

(iii) Behavior segmentation, which involves grouping customers according to how they interact with the product or service o?ered. (iv) Life cycle segmentation or customer journey, apart from understanding buyers' interests and preferences, marketers must also know which buyer is at which stage of the buying process. 3.

Metho ds and results The research method used in this experiment is the CRISP-DM model adjusted to the framework of thought mentioned above, the stages of the CRISP-DM model as shown in Figure 1. Figure 1. CRISP-DM proposed method 3.1. Datasets Data taken from the Martha Tilaar Shop database is related to customer transaction history with customer membership status, for one year there are 15,405 customers. The data cannot be fully used as a dataset in the modeling process using clustering and classi?cation algorithms.

In order to be used as a dataset, the data must pass the data preprocessing stage. Customer pro?le data can be seen in table 1 and customer transaction history data can be seen in table 2 below. The data will go through the data preprocessing stage using the following provisions: (i) Selection of customer age between 15 to 80 years, after that the age in the category into two categories. Age above 50 years with the category "Tua" and age below 50 years with the category "Muda".

(ii) Customer transaction history data selection is only for the number of transactions that are more than 2 for one year. The data that has been carried out in the preprocessing stage occurs to shrink the amount of data to 3,721 data which will then become a dataset used for the RFM analysis modeling process using the Elbow method and the K-Means algorithm. Table 1. Customer Pro?le Information.

ID Customer Age Location Celular 2204130008 Muda Jakarta XL Axiata 2204140015 Tua Jakarta Telkomsel 1111160096 Muda Tangerang Indosat Table 2. Customer Transaction

History. ID Customer Last Trans Value Trans Count Trans 0101130001 2018-03-01 1.535.200 2 0101130007 2018-01-26 331.050 1 0308150019 2018-12-01 1.106.300 3 3.2. Data Clustering At this stage, the research will conduct a process of data clustering.

The data clustering process in this study uses the Elbow method and the K-Means clustering algorithm. The Elbow method in this study uses Python tools to get the desired results. In the Elbow method, a dataset of 3,721 data will be used and an iteration will be started from the value of K = 1 to the value of K = 15.

The results of the use of the Elbow method produced the most optimal number of K is 4. SSE values in each iteration performed by the Elbow method can be seen in Figure 2 below. Figure 2. Graph SSE Elbow Method The clustering stage will be carried out using the K-Means algorithm using cluster number 4, the clustering process uses Rapidminer 9.3

data mining analysis tools. The results of the clustering phase of customer transaction history data with the 3,721 datasets, modeling with K-Means produce clusters with the highest number of members, which are in Cluster 0. Cluster 0 has a total of 2,438 members, as shown in Table 3 below. Table 3. Clustering Process Results.

Cluster Count of Members Cluster 0 2.438 Cluster 1 231 Cluster 2 64 Cluster 3 988 Total 3.721 3.3. Modeling K-Means and Na<sup>¬</sup>ive Bayes This stage will use a customer pro?le information dataset, i.e. ALC, which will be combined with customer data from the pre-formed cluster with K-Means. The process at this stage uses Rapidminer 9.3 data mining analysis tools.

In addition to carrying out the classi?cation process, this stage also carries out a validation process using 10-fold Cross-validation to test the Na<sup>"</sup>ive Bayes classi?cation model. The process of testing the K-Means and Na<sup>"</sup>ive Bayes models. The results of modeling tests can also produce weight values from each ALC attribute in each cluster. The weight value of the ALC attribute in each cluster can be seen in Table 4 below. Table 4. ALC attribute value weights with K-Means and Na<sup>"</sup>ive Bayes.

Atributes Value of Parameter Cluster 0 Cluster 1 Cluster 2 Cluster 3 Usia Muda 0,78 0,75 0,78 0,76 Selular Telkomsel 0,61 0,67 0,64 0,63 Lokasi Jakarta 0,52 0,50 0,56 0,49 3.4. Modeling K-Means and Decision Tree Next is the classi?cation stage of clustering results with K-Means for customer pro?le attributes using the Decision Tree algorithm.

This stage is no di?erent from the classi?cation stage using Na<sup>"</sup>ive Bayes. The validation process was carried out using 10-fold Cross-validation. The classi?cation process using

K-Means + Decision Tree. In testing the K-Means + Decision Tree model it also produces cluster prediction information along with the con?dence value for each customer, so it can know the location of the customer on the actual cluster. Table 5 below shows customer information with predictive value and con?dence for each cluster. Table 5.

ALC attribute value weights with K-Means and Na<sup>--</sup>ive Bayes. Con?dence ID Customer Cluster Cluster pr ed cluster 0 cluster 3 cluster 1 cluster 2 0103130007 cluster 3 cluster 0 0,60 0,30 0,06 0,03 0103130878 cluster 1 cluster 0 0,71 0,22 0,05 0,02 0108140028 cluster 2 cluster 0 0,68 0,24 0,07 0,02 4.

Conclusions The results of testing to obtain an e?ective customer segmentation and pro?ling analysis model using clustering and classi?cation algorithms results that the algorithm model that can provide solutions to these needs is the K-Means + Na<sup>"</sup>ive Bayes algorithm model, with a high accuracy value of 65, 87%. In addition, this study found a large number of customer clusters, but not so good to contribute to the revenue of Martha Tilaar Shop, this cluster is in cluster 0.

With the results of this study, it is expected that customers in cluster 0 get the attention of Martha Tilaar Shop to be able to increase and contribute greatly to the company's revenue. Besides that, the most in?uential customer pro?le in the segmentation and Martha Tilaar Shop's customer pro?le is in the Age attribute with the value of "Muda", the Location attribute with the value of "Jakarta" and the Cellular attribute with the value of "Telkomsel".

References [1] Nimbalkar D D, 2013 Data Mining Using RFM Analysis Int. J. Sci. Eng. Res. 4, 12 p. 940–943. [2] Niyagas W Srivihok A and Kitisin S, 2006 Clustering e-Banking Customer using Data Mining and Marketing Segmentation ECTI Trans. Comput. Inf. Technol. 2, 1 p. 63–69. [3] Malang B, 2017, Pentingnya Mengelola Hubungan dengan Pelanggan dalam Bisnis, http://binus.ac.id/. [On- line]. Available: http://binus.ac.id/malang/2017/09/pentingnya-mengelola-hubungan-dengan-pelangga n- dalam-bisnis/. [Accessed: 15-May-2019]. [4] University B, 2017, Customer Relationship Management, https://sis.binus.ac.id. [Online]. Available: https://sis.binus.ac.id/2017/04/20/customer-relationship-management/.

[Accessed: 15-May-2019]. [5] Roshan H and Afsharinezhad M, 2017 The New Approach in Market Segmentation by Using RFM Model J. Appl. Res. Ind. Eng. 4, 4 p. 259–267. [6] Maryani I and Riana D, 2017 Clustering and pro?ling of customers using RFM for customer relationship management recommendations 2017 5th Int. Conf. Cyber IT Serv. Manag. CITSM 2017 p. 2–7. [7] Rouse M, 2005, RFM analysis (recency, frequency, monetary), https://searchdatamanagement.techtarget.com. [Online].

Available: https://searchdatamanagement.techtarget.com/de?nition/RFM-analysis. [Accessed: 25-Jun-2019]. [8] Nair A, 2018, RFM Analysis For Successful Customer Segmentation, https://www.putler.com. [Online]. Available:

https://www.putler.com/rfm-analysis/. [Accessed: 20-Jun-2019]. [9] Fader P S Hardie B G S and Lee K L, 2005 RFM and CLV: Using Iso-Value Curves for Customer Base Analysis J. Mark. Res. 42, 4 p. 415–430.

[10] Kotler P and Armstrong G, 2004 Dasar-dasar Pemasaran, Jilid 1 9th ed. Jakarta: Indeks, Jakarta. [11] Meitasari S, 2018, Segementasi Pelanggan – Kunci Sukses Dalam Bisnis, https://disperin.ntbprov.go.id. [Online]. Available:

https://disperin.ntbprov.go.id/content/segmentasi-pelanggan-—kunci-sukses-dalambisnis. [Accessed: 23-May-2019]. [12] Silalahi M, 2018, 4 Tipe Segmentasi Pelanggan, www.mix.co.id. [Online]. Available:

https://mix.co.id/marcomm/news-trend/4-tipe-segmentasi-pelanggan/.

[Accessed: 04-Jul-2019].

## INTERNET SOURCES:

\_\_\_\_\_

<1% -

https://www.mbaknol.com/modern-management-concepts/business-continuity-management/

<1% - https://publications.waset.org/vol/64

<1% -

https://www.researchgate.net/publication/336890527\_Customer\_Relationship\_Management\_and\_Customer\_Retention

<1% - http://docshare.tips/japan-christian-year-book\_574b1d83b6d87fa21f8b4de4.html <1% -

https://hbswk.hbs.edu/item/how-do-you-weigh-strategy-execution-and-culture-in-an-organizations-success

<1% -

https://www.marketing-schools.org/types-of-marketing/relationship-marketing.html <1% - https://patents.google.com/patent/US20100311496A1/en

<1% -

https://www.entrepreneurshipinabox.com/12312/5-things-when-your-customers-stop-buying/

1% -

https://www.dwsassociates.com/marketing-tools/multi-channel-marketing-tools/rfm-sc

oring-calculator/

1% - https://www.putler.com/rfm-analysis/

<1% - https://en.wikipedia.org/wiki/RFM\_%28customer\_value%29

<1% -

https://www.researchgate.net/publication/307745428\_Strategic\_market\_segmentation

<1% - https://www.questionpro.com/blog/demographic-survey-questions/

<1% - https://instapage.com/blog/demographic-segmentation

<1% -

https://www.slideshare.net/nagababungr/consumer-behavior-and-marketing-strategiesunit1-and-unit-2

<1% -

https://id.123dok.com/document/zpd89l0z-big-data-optimization-recent-developments -and-challenges-pdf-pdf.html

<1% -

http://openaccess.thecvf.com/content\_CVPR\_2020/papers/Wang\_Dual\_Super-Resolution \_Learning\_for\_Semantic\_Segmentation\_CVPR\_2020\_paper.pdf

<1% - https://issuu.com/asdfjrnl/docs/978-81-929866-1-6

<1% - https://jakevdp.github.io/PythonDataScienceHandbook/05.11-k-means.html <1% -

https://subscription.packtpub.com/video/big\_data\_and\_business\_intelligence/97817889 94989/44632/44639/k-means-clustering

<1% -

https://www.coursera.org/lecture/machine-learning-data-analysis/running-a-k-means-cl uster-analysis-in-sas-pt-1-hVbQ3

<1% -

https://www.researchgate.net/publication/313554124\_Review\_on\_Determining\_of\_Cluste r\_in\_K-means\_Clustering

<1% -

https://www.researchgate.net/publication/318056374\_Comparison\_of\_Naive\_Bayes\_Ran dom\_Forest\_Decision\_Tree\_Support\_Vector\_Machines\_and\_Logistic\_Regression\_Classifier s\_for\_Text\_Reviews\_Classification

<1% -

https://www.city.komae.tokyo.jp/index.cfm/50,96080,c,html/96080/20181013-184705.pd f

<1% - https://arxiv.org/pdf/1803.02816.pdf

1% - https://link.springer.com/article/10.1007/s10257-014-0267-y

<1% - https://link.springer.com/article/10.1007/s11301-012-0088-3

<1% - http://scholar.unand.ac.id/15224/4/Daftar%20Pustaka\_1.pdf