Customer Loyalty Analysis Using Customer Lifetime Value (A Case Study of Baby Equipment SMEs)

M. Dachyar and Lidya Vitasya
Industrial Engineering Department
Universitas Indonesia
Depok, Indonesia
mdachyar@yahoo.com, lidyavitasya25@gmail.com

Abstract

The increase number of births in Indonesia during the pandemic era provided opportunities for baby equipment businesses. At the same time, online purchase transactions have also increased. This research objective is to determine the segmentation and worth of each customer segment using Customer Lifetime Value (CLV) and develop strategies for every customer segmentation in order to increase the value of business competitiveness. The K-Means Clustering method is applied to segment clients into a few clusters and the CLV value is used to specify the value of each customer segment with the RFM variable. Cluster deployment using Customer Value Matrix (CVM) is also carried out to ensure cluster characteristics. Secondary data is obtained from sales transactions. Analysis of product sales is carried out using the Association Rules method that results in one of the strategies in Customer Development. The research resulted in 5 clusters for customers with 16 strategies for the whole cluster. Cross-selling strategy is the most recommended strategy.

Keywords
Customer Lifetime Value (CLV), Business Intelligence, K-Means Clustering, Customer Value Matrix, Association Rule

1. Introduction

Pandemic in Indonesia has provided opportunities for business in the baby equipment sector because birth rate increased by 15%, also online shopping in e-commerce increased 25% in sales (“IdEA: Kenaikan Penjualan E-Commerce 25 Persen Selama Pandemi - Bisnis Tempo.Co,” n.d.). This situation gives opportunity for business owners who focused on baby equipment sector to expand their market through e-commerce platforms. Customer Relationship Management (CRM) is used to build and maintain the relationship between business and its customers. It is important for businesses to use CRM and maintain their customers, especially those that are profitable for the company(Dachyar, Esperanca, & Nurcahyo, 2019). CRM is important for businesses to know how to find out the loyalty of the customers that they have.

1.1 Objectives

This study aims to find out customer segmentation on baby equipment business. By knowing the segmentation of the customers, business can count the value of CLV on each of the segments so that strategies for each segmentation could be formed to help business increase its competitiveness value.

2. Literature Review

E-commerce provides amenities during the time spent computerized exchange that goes for online merchants and customers (Nisar & Prabhakar, 2017). E-commerce is a tool that fasten the business transaction process through the internet that includes the exchange of valuable information in products and services form, also payment with a technology that is web-based (H. Zaied, 2012). Based on the relationship of sellers and buyers, e-commerce can be categorized in five types, which are: B2B Commerce (transaction between two business units), B2C Commerce (a transaction between business unit to personal customers), C2B Commerce (transaction where the individual is the seller and the business unit is the buyer), C2C Commerce (transaction between individuals), and Cooperative Trade (all the colleague in the business inventory network are collaborating) (Turban, King, Lee, Liang, & Turban, 2015). There are so many benefits customers could have through e-commerce, such as the increase of products variety, less operational cost, and interactivity (Nisar & Prabhakar, 2017). But this also comes with risks for the customers. Some of the risks are: fake web, fraud, security, and unable to know the real condition and quality about the products that are bought (Pratima Bhalekar, 2014).

CRM is a managerial activity that manage all interaction with customers in a business by combining business process and technology to help understanding customers wholly (Dachyar & Hananto, 2014). CRM has these characteristics, which are relationship management, an automatic salesforce, the use of technology, and opportunities management (Zeng, Li, & Ding, 2013). There are also three types of CRM, which are strategic CRM (focusing to win and maintain customers), operational CRM (focusing on process automation that corresponds with customers), analytic CRM (business changes data about customer in order to find knowledge that are going to be put into strategic or tactic objectives) (Buttle & Maklan, 2015). The objectives of CRM is to create a long-term relationship with beneficial customers.

Customer Lifetime Value (CLV) is the current worth of the multitude of net edges the organization got from client connections and division (Buttle & Maklan, 2015). This customer relationship is seen as a decision to invest and customers as the generator incomes. Since maintaining relationships with customers requires the company to spend money to keep them, by using CLV, the company can measure all the income and spending that are related to customer relationships.

Association rules or known as affinity analysis is an examination about "what with what". For example, in a transaction study on supermarkets, a customer wanted to buy soap and toothpaste at the same time. These association rules want to give the information in an "if-then" relationship. This association is the sum of probabilistic data. The result of these association rules can help with a lot of decision making in the company, such as catalog design, cross-marketing/cross-selling, and loss-leader analysis (Sumathi & Esakkirajan, 2007). The purchase patterns show that there is a similarity in buying behaviors from the customers with the same RFM value and demographic variable (Birant, 2011).

Business Intelligence (BI) is a blend of equipments, strategies, and cycles that assists the organization with augmenting the information that they had so it very well may be another information for settling on a superior choice (Jennex & Bartczak, 2013). BI lets company to use their data as best as possible. BI system combines data collection, storing, and operating the knowledge and information flow with tools to analyze internal and complex information for the companies. This will help the company to answer questions in operating the business.

3. Methods

This study used one transaction data-set from 1 unit business which is selling baby equipment such as upper clothes, lower clothes, and other baby products in Indonesia. The data-set contains the customer ID, product ID, transaction date, the quantity of order, the amount of billing, and customer location that contains data transaction in 12 months or approximately 1 year.

Transaction and products data that have been collected would be sorted out to get only useful and important information for being used as the basis to make decisions on the company. Spotfire Analytics is used for database processing. In this step, we chose, collected, and changed data, and the result of this step is values of every RFM variable for every customer that will be used to analyze the customer segmentation.

RFM variables that composed of recency, frequency, and monetary were collected during the research process (Bult & Wansbeek, 1995). Recency is the interval of the last transaction to the last research date. Frequency is the number of times transaction happened. Monetary is the total of money that customer spent to the company during the research period. The RFM variables can be used to process K-Means Algorithm (Hosseini, Maleki, & Gholamian, 2010) and determine CRM in company (Wu et al., 2020). This research period went from March 2020 until March 2021

K-Means algorithm is a clustering technique that is used with RFM variables. This method is considered efficient even with repeated experiments (Tan & Steinbach, 2006). With this method, cluster number depends on the decision maker. Steps on using K-Means are: found the first cluster number, then allocated to the centroid with the closest distance, and recount again with the new centroid so that there will be no data that changes cluster (Zhu, Wang, Wu, & Zhu, 2011). Clustering with K-Means was done with Spotfire Analytics. The Euclidean Distances method was used as the basis to cluster the data with existing RFM attributes. This method will generate a cluster with the maximum distance between cluster and minimum distance between each member in the cluster (Zhu et al., 2011).

Customer loyalty was determined by the rank of CLV on each customer segmentation from the last step. A few CLV models came from the basic equations (Buttle & Maklan, 2015). One of those is the equation used to analyze customers in the the e-commerce industry and fashion/beauty companies. The equation will be written below. (Khajvand, Zolfaghar, Ashoori, & Alizadeh, 2011; Liu & Shih, 2005).

$$CLV = (NR \times \text{Weight }) + (NF \times \text{Weight } F) + (NM \times \text{Weight } M) \qquad (1)$$

The first thing that should be done is to normalize the value of each RFM variable because they have a different range. The Min-Max normalization method was used to normalize the value.

$$\text{NV} = \frac{\text{V} - \text{Min}}{\text{Max} - \text{Min}} New_{max} - New_{min} + New_{min} \qquad (2)$$

After normalizing the RFM variables value, we calculated the weight of RFM variables using pairwise comparison methods with the Analytical Hierarchy Process (AHP) that needed experts assessment by using a questionnaire. The results then were processed by using Expert Choice to show the weight of each variable which presented the experts' preferences. Steps of processing AHP are:

1. Value of importance would be done by pairing two attributes with AHP scale, which is 1-9.
2. Calculated the inconsistent of the experts' preferences by using pairwise comparison. The value of inconsistency should not be more than 10% or 0.1.
3. Combined every result of the experts' preferences and calculated the weight of each attribute.

CVM was used to see the distribution of each cluster. Mapping with matrix could help company to understand the result of segmentation better as the basis of decision making (Marcus, 1998). CVM uses the average value of the two highest attributes based on the results of the previous step from each cluster as a separator between the quadrants formed.

Association Rules Mining identifies associations between a set of product items that are oftenly bought together by analyzing customer purchases of a product (Agrawal, Imielinski, & Swami, 1993). Processing association rules used STATISTICA software. We are free to determine the minimum support and minimum confidence values according to our needs (Loughin et al., 2008).

4. Data Collection

The data collection is the data-set that has the transactions and products information during the research period. The summary of data transaction that has been collected will be presented on the Table 1

Table 1 Summary of Transaction Data

Data Type	Total
Total Transaction Data	18901 transaction
Total of Transactions ($)	$605615,38
Number of Customers	6873 customers
First Payment Date in Specified Period	01 March 2020
Last Payment Date in Specified Period	31 March 2021

In Table 1, we can see that the total of the data transaction is 18901 with 6873 customers during research period. The total amount of transaction that has been done equals to $605615,38 which categorize this into the medium size SMEs. Summary of products information is presented on Table 2 below. Most products this business had is categorized into other products. The total of items that this business had is 123 items.

Table 2 Product Data Summary

Data Type	Total
Upper Clothe (Code: B)	20 products
Lower Clothe (Code: C)	24 products
Other Products (Code: PL)	79 products

5. Results and Discussion

All the datas that have been processed before this step was processed by using Spotfire Analytics. Data transaction consists of customers ID, total amount, currency, date of transaction, location, and lists of items that customer purchased. Here is some of the transaction data used in this study is represented in Figure 1 below.

1	No	Nomor Pelanggan	Total Amount	Currency	Tanggal Transaksi	Lokasi Pengiriman	Baju	Baju 2	Baju 3	Baju 4	Celana
2	1	NP3792	507386	IDR	3/1/2020	Medan	B17	B18			C17
3	2	NP2628	688285	IDR	3/1/2020	Medan	B13	B12			C11
4	3	NP3485	343637	IDR	3/1/2020	Gorontalo					C20
5	4	NP4817	1028419	IDR	3/1/2020	Medan	B19	B18			C22
6	5	NP5788	662182	IDR	3/1/2020	Tebing Tinggi	B2	B14			C2
7	6	NP3513	533079	IDR	3/1/2020	Tebing Tinggi	B9	B8			C10
8	7	NP5962	550315	IDR	3/1/2020	Siantar	B3				C11
9	8	NP6794	784364	IDR	3/1/2020	Medan	B7	B9			C17
10	9	NP5169	688086	IDR	3/1/2020	Medan	B5	B2			C24
11	10	NP384	294422	IDR	3/1/2020	Jakarta	B12	B7			C22
12	11	NP3473	1149772	IDR	3/1/2020	Tanggerang	B17	B14			C5
13	12	NP752	511202	IDR	3/1/2020	Siantar	B11	B7			C3
14	13	NP2478	1096416	IDR	3/1/2020	Gorontalo	B20	B18			C18
15	14	NP3035	1167596	IDR	3/1/2020	Medan	B12	B16			C23
16	15	NP4584	1125589	IDR	3/2/2020	Jakarta	B19	B2			C2
17						:					
18						:					
19	18891	NP3611	315842	IDR	3/31/2021	Aceh	B10				
20	18892	NP4957	34410	IDR	3/31/2021	Tanggerang	B12				
21	18893	NP5540	104613	IDR	3/31/2021	Siantar	B8				
22	18894	NP6868	193647	IDR	3/31/2021	Jakarta	B14				
23	18895	NP6872	363438	IDR	3/31/2021	Tanggerang	B1				C6
24	18896	NP6867	139704	IDR	3/31/2021	Aceh	B14				C21
25	18897	NP6871	653266	IDR	3/31/2021	Tanggerang	B20				C13
26	18898	NP6873	713854	IDR	3/31/2021	Gorontalo	B19				C3
27	18899	NP6872	344318	IDR	3/31/2021	Tanggerang	B12				C24
28	18900	NP6869	887129	IDR	3/31/2021	Medan	B2				C2
29	18901	NP6867	99721	IDR	3/31/2021	Siantar	B5				C9
30	18902	NP6867	387884	IDR	3/31/2021	Tanggerang	B17				C1

Figure 1. Some of Transaction Data

Top 10 most purchase items during this research times will be presented in Figure 2. This is the result of processing data with Spotfire Analytics. In Figure 2, it can be seen that products are from Code B and C. The most purchase item is B7 with the total of 2181 times.

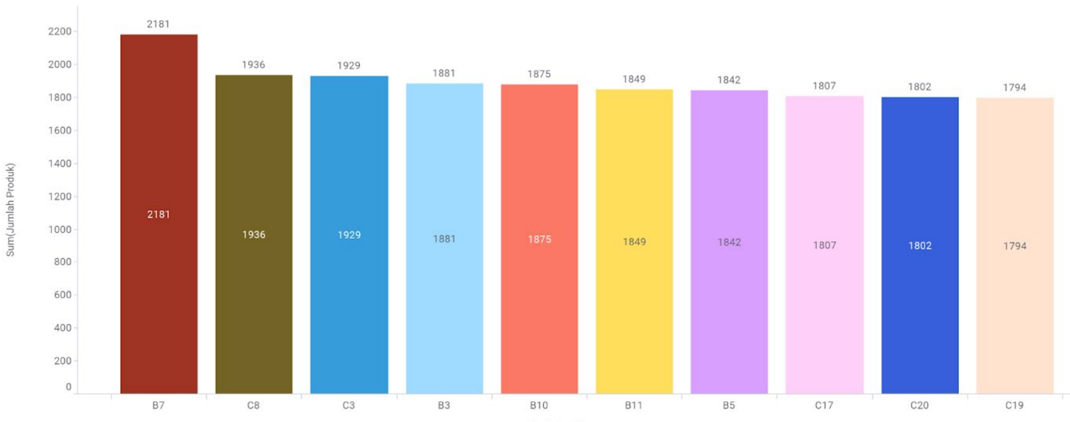

Figure 2. Top 10 Most Purchase Items

Figure 3 below will point out the number of total transaction for every location that is written in the data. This is processed by Spotfire Analytics. By the Figure 3 below, it can be seen that Medan, the location of the store has the most transaction during the research period, taking up into 30.95% from the total transaction that has been collected. This is followed by Siantar, with the total transaction of 3018, taking up into 15.97% from the total transaction.

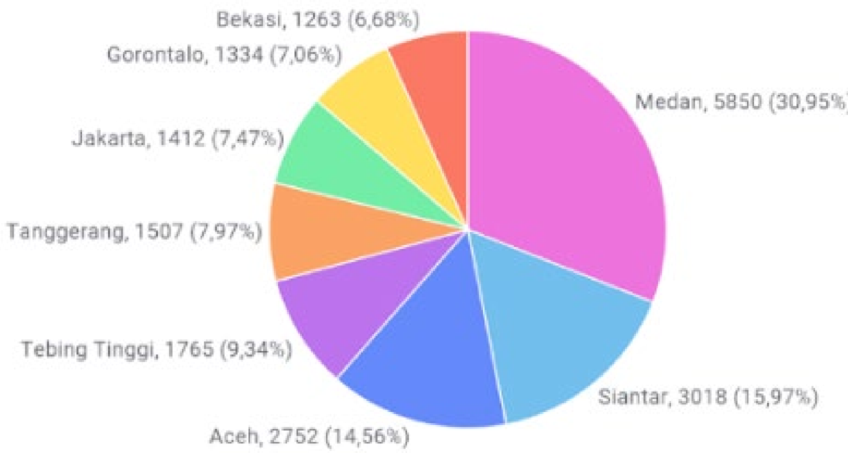

Figure 3. Transactions per Location

The data contained in Figure 4 is processed in Spotfire Analytics and has been processed into a database with RFM variables for each customer. Database has been filtered to important data that will be useful for the research. The data from the RFM variable filtering resulted in a total of 6873 according to the number of existing customers. This RFM attribute filtering was done by inputting transaction data, totaling 18901 data. The results of this RFM attribute indicated that most of the company's customers had repeated transactions or re-purchases.

Nomor Pe...	Sum(Recency)	Sum(Frequen...	Sum(Monetary)
NP1	278,00	1,00	645714,00
NP2	33,00	3,00	548942,00
NP3	3,00	5,00	1940399,00
NP4	12,00	1,00	1068192,00
NP5	196,00	2,00	642230,00
NP6	73,00	2,00	634761,00
NP7	23,00	4,00	1841076,00
NP8	34,00	5,00	2214507,00
NP9	3,00	3,00	1674403,00
NP10	2,00	4,00	1903278,00

Figure 4. Part of the result of RFM's value to the customer

K-Means Clustering was carried out with the number of K (clusters) = 3, 4, 5, and 6 using the Euclidian Distance method. As can be seen in Table 3, that after looking at the training error of each cluster, it was found that cluster 5 had the smallest training error value. Number of cluster was chosen by the least value of training error. The value of training error was generated using Spotfire Analytics.

Table 3. Number of Training Errors in Each Cluster (Masukin cluster 6)

Number of Cluster	Training Error
3	1,008346

Number of Cluster	Training Error
4	1,008112
5	1,005339
6	1,006152

Normalization of each RFM variable for each cluster is performed. RFM variable value can be seen in Table 4 below. Normalization of each RFM variable is important because each of the variable has different range. (Tulis tujuan untuk melakukan normalisasi -> untuk menghitung CLV)

Table 4. Normal Value of Each RFM Variabel Variable

Cluster	NR	NF	NM
1	0,68	0,05	0,04
2	0,79	0,1	0,12
3	0,85	0,15	0,21
4	0,91	0,25	0,33
5	0,99	0,6	0,71

Weight calculation was done for each of the RFM variables, which was required to determine this questionnaire distributed to the experts. Table 5 will display the summary results of the expert's assessment of the weight of the RFM variable. Table 5 shows that three out of four experts feel that the frequency variable is the most important variable when compared to other RFM variables. It's because experts said that it is better to have customers with a very high purchase intensity compared to a large number of purchases but only one time. A high-frequency level may reflect a high level of loyalty and a good relationship between customer and company-owned.

Table 5. Summary of Expert Assessment of the RFM Variable Thickness

Experts	Recency	Frequency	Monetary
1	3	1	2
2	3	1	2
3	3	1	2
4	3	2	1

Weight of RFM variables is determined by combining all the experts' assessment. Table 6 below will show the results of combining all expert assessment from every attributes of RFM. Frequency has the highest weight between the attributes, followed by monetary and recency.

Table 6. Weight of Reviewer Variables, Frequency, Recency, and Monetary

Attributes	Weight
Recency	0,132
Frequency	0,575
Monetary	0,293
Total	1

CLV is obtained by using normalized value and weight of variables by using equation (1). Table 7 below will show the CLV value and ranking of each cluster, where cluster 5 has the highest CLV value, thus getting a rating of 1. High CLV value can be categorized into clusters with more loyal customers.

Table 7. CLV Value and Rank in Each Cluster

Cluster	CLV	CLV Rank
1	0,13	1
2	0,20	2

Cluster	CLV	CLV Rank
3	0,26	3
4	0,36	4
5	0,68	5

Table 7 shows that cluster 1 has the lowest rank, while cluster 5 has the highest rank. In addition to know the order of customer loyalty in each cluster, this ranking will help companies to analyze strategies to maintain customer loyalty. (Liu & Shih, 2005).

Customer value is be attributed to the analysis that has been done before, by analyzing clusters of customers based on the model RFM (type/characteristics of customers). The analysis of this section will see whether the characteristics of the defined RFM variables have the appropriate CLV values (quantitatively) for all the characteristics of the RFM variables (Birant, 2011; Liu & Shih, 2005). RFM combination on each cluster will be shown in Table 8 below.

Table 8. Combination of CLV Calculation and RFM Characteristics

Cluster	Number of Customers	CLV Rank	RFM Combination	Customers Characteristic
1	2176	5	R↓F↓M↓	*Uncertain Customer*
2	2262	4	R↓F↓M↓	*Uncertain Customer*
3	1482	3	R↑F↓M↓	*First Timer Customer*
4	529	2	R↑F↑M↑	*Best / Valuable Customer*
5	4	1	R↑F↑M↑	*Best / Valuable Customer*

Segment mapping based on CVM is needed because there are still shortcomings in the segmentation results using the K-Means method only. Mapping with this matrix is used effectively to help to understand more deeply the results of segmentation as the basis for making a decision (Marcus, 1998). CVM uses the average value of the frequency and monetary attributes of each cluster as a separator between the quadrants formed. Table 9 below will show the average value used as a dividing line between quadrants.

Table 9. Average Axis on CVM

Axis X	Total frequency of 5 clusters	28
	Average value of frequency	5,6 ~ 6
Axis Y	Total monetary from 5 clusters	$993.69
	Average value of monetary	$198.74

Each cluster will be mapped into the CVM according to the axis. Figure 5 will show the results of cluster mapping using CVM. Cluster 5 is in best quadrant, cluster 4 is in between best and spender, and cluster 3, 2, 1 are respectively in uncertain quadrant.

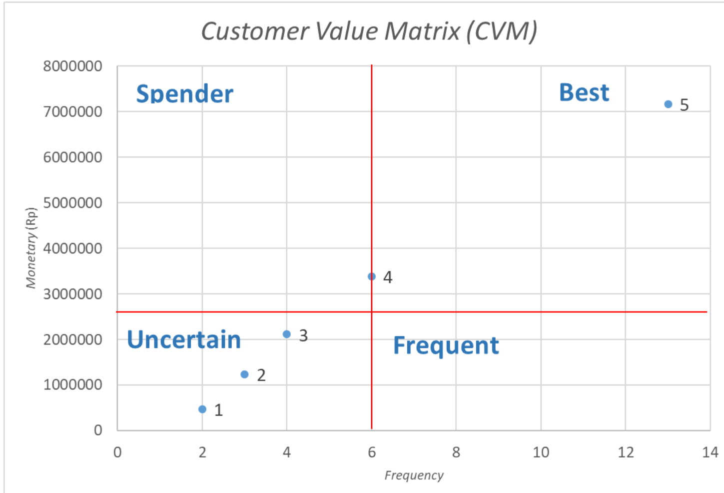

Figure 5. CVM on Customer Clusters

Table 10 will provide a clearer description of the distribution of customers with the result that quadrant 3 is the quadrant with the highest number of clusters, namely three.

Table 10. Mapping Description on CVM

Quadrant	Cluster	Quadrant Characteristic
Quadrant 1	Cluster 5	Best
Quadrant 2	Cluster 4	Spender
Quadrant 3	Cluster 1 Cluster 2 Cluster 3	Uncertain
Quadrant 4	-	Frequent

Product data processing is carried out using Association Rules. Association Rules are used to find the minimum support association rules and requirements minimum confidence. The help worth of the affiliation rule states how frequently the standard is utilized onto the information. A more prominent help esteem implies a solid relationship between's item things. Certainty esteem is a proportion of the dependability of a gathering rules (Agrawal et al., 1993). An itemset is viewed often, if the support value of the itemset surpasses the minimum support value specified by the user. Association rules that satisfy a specialist characterized least certainty worth could be resolved from the itemset..

The data entered is also a transaction that has 2 product purchases in one transaction only. If a transaction has more than 2 products, the third product and so on are not used. This is done to maximize the number of transactions. Table 11 will display a summary of the product data that has been owned.

Table 11. Product Data Summary

Data Type	Total
Total product data used	123 products
Number of transactions	18692 transactions
Date of first payment within the specified period	1 March 2020
Last payment date within the specified period	31 March 2021

We are free to determine the minimum support and minimum confidence values according to our needs (Loughin et al., 2008). For example, if you want to find data that has a fairly strong association relationship, the minimum support and minimum confidence can be given a fairly high value. On the other hand, if we only want to see the number of variations in the data without being too concerned about the strength or weakness of the association between the data, the minimum value can be set low.

Results of the determination of this minimum value will affect the results of the final process of association rule, and the following is the process of association rule that is done to obtain the optimal product bundling by using STATISTICA software. Here are the steps in doing association rules:

1. Entering a database that has been created entirely in the STATISTCIA software, with the headers in the data used as variables in the software.
2. Entering the variables that will be analyzed in the next process
3. Determining the value of minimum support, minimum confidence value, and maximum item sets in body and maximum item set in head that will determine the outcome of treatment.
4. Repeat the third step if you want to get association results that have more relationship strength than those generated previously by increasing the minimum support and confidence values entered.

Minimum value of support and confidence put by the researcher is 0.2%. The results of this process show 6 rule sets which will be shown in Table 12 below. Confidence value ranges from 0.36%-0.68%. Support value ranges from 0.2%-0.23%. In Table 12 it can be seen that the largest confidence value is in rule number 1, namely C10 -> B3 with a support value of 0.21% and a confidence value of 6.8%. The support value of 0.21% means that this rule has appeared 40 times for purchases. The confidence value obtained alone is 6.8%, which means that if in a transaction there is B10, then the probability of buying a B3 product is 6.8% from 40 times, which is 3 times.

Table 12. Association Rules Result

No	Body	==>	Head	Support(%)	Confidence(%)
1	C10	==>	B3	0.21	0.68
2	C5	==>	B7	0.23	0.68
3	C7	==>	B20	0.20	0.63
4	B3	==>	C10	0.21	0.39
5	B7	==>	C5	0.23	0.39
6	B20	==>	C7	0.20	0.36

The characteristics of each cluster are generated which are used to produce customer development strategies, which are carried out to maintain relationships with customers so that customers can continue to contribute to business development, and can improve the company's competitiveness (Khajvand et al., 2011).

From the results of the analysis of customer segmentation and distribution on the CVM matrix, clusters 4 and 5 have the highest loyalty, then the first thing to do is to create a loyalty point feature on customers' accounts. (Peker et al., 2017). Furthermore, further developing assistance as far as conveyance, giving quicker data about items can likewise keep up with the connection between the organization and its clients (Khajvand et al., 2011; Valvi & Fragkos, 2012)

In the cluster that is included in the first time and uncertain customer in CLV analysis using RFM and enters the third quadrant with the uncertain category on the CVM matrix, namely clusters 1,2, and 3, increasing customer awareness of the company and customers can maintain and improve relationships with customers that will make a growing desire to transact (Buttle & Maklan, 2015; Wang & Singh, 2006). Special services provided online and in real-time can affect customer loyalty (Dachyar & Athory, 2015) The use of social media to deepen customer engagement and delivery of questionnaires to determine the pattern and consideration to purchasing customers can help companies to develop strategies for further (Peker, Kocyigit, & Eren, 2017).

Cross-selling and up-selling methods are also methods that can be added in the implementation of customer development strategies to lure customers to increase the number of purchases (Buttle & Maklan, 2015). Recommendations for product equivalents for cross-selling are also found in the results of product data processing using Association Rules.

6. Conclusion

This study intends to help the baby equipment business in investigating customer loyalty based on CLV segmentation. The outcomes called attention to five customer clusters based on RFM characteristics, namely Best/Valuable, First

Timer, and Uncertain customers. There are also tactical steps that can be used by businesses to improve and maintain customer relationships with the company to give an impact on customer loyalty. These steps are based on RFM characteristics by preserving customer convenience, improving customer trust in the company, attracting customers to make repeat transactions, and increasing customer satisfaction and customer service quality.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93*, 207–216.

Birant, D. (2011). Data Mining Using RFM Analysis. *Knowledge-Oriented Applications in Data Mining*. https://doi.org/10.5772/13683

Bult, J. R., & Wansbeek, T. (1995). Optimal Selection for Direct Mail. *Marketing Science*, *14*, 378–394.

Buttle, F., & Maklan, S. (2015). *Customer Relationship Management, Concept and Technologies, Third Edition*.

Dachyar, M., & Athory, E. S. (2015). Improving the performance of customer loyalty of online ticketing in Indonesia's showbiz industry. *Journal of Physics: Conference Series*, *622*. https://doi.org/10.1088/1742-6596/622/1/012016

Dachyar, M., Esperanca, F. M., & Nurcahyo, R. (2019). Loyalty Improvement of Indonesian Local Brand Fashion Customer Based on Customer Lifetime Value (CLV) Segmentation. *IOP Conference Series: Materials Science and Engineering*, *598*. https://doi.org/10.1088/1757-899X/598/1/012116

Dachyar, M., & Hananto, L. (2014). Innovation and quality service factors to customer loyalty in indonesia telecommunication company by using structural equation modeling method. *Management and Technology in Knowledge, Service, Tourism and Hospitality*, 41–44.

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*, 5259–5264.

H. Zaied, A. N. (2012). Barriers to E-Commerce Adoption in Egyptian SMEs. *International Journal of Information Engineering and Electronic Business*, *4*, 9–18.

IdEA: Kenaikan Penjualan E-commerce 25 Persen selama Pandemi - Bisnis Tempo.co. (n.d.). Retrieved March 9, 2021, from https://bisnis.tempo.co/read/1404513/idea-kenaikan-penjualan-e-commerce-25-persen-selama-pandemi?page_num=2

Jennex, M. E., & Bartczak, S. E. (2013). A revised knowledge pyramid. *International Journal of Knowledge Management*, *9*, 19–30.

Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, *3*, 57–63.

Liu, D. R., & Shih, Y. Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information and Management*, *42*, 387–400.

Loughin, T. M., Austin, M., Hand, D. J., Drake, J. M., Holford, N., Behr, S., … Loughin, T. M. (2008). *Data Mining: Methods and Models*. 309–319.

Marcus, C. (1998). A practical yet meaningful approach to customer segmentation. *Journal of Consumer Marketing*, *15*, 494–504.

Nisar, T. M., & Prabhakar, G. (2017). What factors determine e-satisfaction and consumer spending in e-commerce retailing? *Journal of Retailing and Consumer Services*, *39*, 135–144.

Peker, S., Kocyigit, A., & Eren, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: a case study. *Marketing Intelligence and Planning*, *35*, 544–559.

Pratima Bhalekar, S. I. K. P. (2014). the Study of E-Commerce. *Asian Journal of Computer Science and Information Technology*, *4*, 25–27.

Sumathi, & Esakkirajan. (2007). *Fundamentals of Relational Database Management System*.

Tan, P., & Steinbach, M. (2006). Introduction to Data Mining Instructor ' s Solution Manual. *Names*, *28*, 9–35, v.

Valvi, A. C., & Fragkos, K. C. (2012). Critical review of the e-loyalty literature: A purchase-centred framework. *Electronic Commerce Research*, *12*, 331–378.

Wang, Y., & Singh, M. P. (2006). Trust representation and aggregation in a distributed agent system. *Proceedings of the National Conference on Artificial Intelligence*, *2*, 1425–1430.

Wu, J., Shi, L., Lin, W. P., Tsai, S. B., Li, Y., Yang, L., & Xu, G. (2020). An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K -Means Algorithm. *Mathematical Problems in Engineering*, *2020*. https://doi.org/10.1155/2020/8884227

Zeng, F. J., Li, H., & Ding, L. (2013). Database design of merchandise management system based on SQL server. *Applied Mechanics and Materials*, *380–384*, 4728–4731.

Zhu, T., Wang, B., Wu, B., & Zhu, C. (2011). Role defining using behavior-based clustering in telecommunication network. *Expert Systems with Applications*, *38*, 3902–3908.

Biography

M. Dachyar is a Professor and Head of MISDS Laboratory in Industrial Engineering Dept. at Universitas Indonesia. His research is centered around the management of information system, the decision support system, business intelligence, business process reengineering, and operations management.

Lidya Vitasya is a research associate of MISDS Laboratory in Industrial Engineering Dept. at Universitas Indonesia. Her research is centered around the customer relationship management and customer strategic.