© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model

Fitri Marisa[1,2*], Sarifah Shakinah Syed Ahmad[2], Zeratul Izzah Mohd Yusof [2,] Fachrudin[1], Tubagus Mohammad Akhriza Aziz[3]

[1] Department of Informatics Engineering, Widyagama University of Malang, Jalan Borobudur no 35, Malang, 65125, INDONESIA

[2] Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Jalan Hang Tuah jaya Durian Tunggal, Melaka, 76100, MALAYSIA

[3] Department of Informatics Engineering, Pradnya Paramita School of Informatics Management and Computer, Jalan LA Sucipto no 249-A, Malang, INDONESIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.00.00.0000.00.0000
Received 24 July 2019; Accepted 31 July 2019; Available online 3 September 2019

Abstract: The CLV model is a measure of customer profit for a company that can be used to evaluate the future value of a customer. The CLV model is a measure of customer profit for a company that can be used to evaluate the future value of a customer. This study aims to obtain Customer Lifetime Value (CLV) in each customer segment. Grouping uses the K-Means Clustering method based on the LRFM model (Length, Recency, Frequency, Monetary). The cluster formation process uses the Elbow Method and SSE with the best number of clusters = 2 clusters. CLV values are generated from the multiplication of the results of normalization of LRFM and the LFRM weight values are then summed, and carried out on each cluster that has been formed. The highest ranking among the 2 clusters is at the second cluster with the CLV value being far the highest from the other cluster average of 0.362. Based on LRFM matrix, this cluster has a high loyalty value with the symbol LRFM L↑ R↑ F↑ M↑ which is a loyal customer (the best segment that has high customer loyalty value). Based on the LRFM symbol, the company can make a strategy to retain customers and acquire customers to become loyal customers with high profitability.

Keywords: CLV, LRFM, K-Means, Clustering, SMEs

1. Introduction

In this decade, SMEs have experienced substantial growth. According to the results of research conducted by the Retail Research Center, this sector experienced a growth rate of 18.6% in Europe in 2015 and 16.7% in 2016. The increasing competition in the SMEs demanded this effort to improve techniques and strategies to maintain customer satisfaction levels to continue to increase .[1]. The SMEs sector has an important role in the country's economy, especially Indonesia. They have proven their existence in the past few years. SMEs have a proportion of 99.99% of the total business actors in Indonesia or as many as 56.54 million units. Based on data achieved by SMEs, in 2013 the

*Corresponding author: fitrimarisa@widyagama.ac.id

value of exports increased by 9.29% or Rp.182 billion, of which almost 86.33% used independent capital. [2]. Therefore, efforts are needed to improve the progress of SMEs.

The customer is an asset for the company, and they have different preferences.[3]. At present, companies are competing to attract the attention of customers, and this reseach [4] is one of field has studied it. In reality, each customer has different needs, expectations and behaviors, so the company is expected not to treat them with the same treatment. The Customer not only have different preferences, expectations, and needs but also different income and profiles of cost, therefore must be managed in different treatments. [5],[6],[7]. The same treatment for all customers will cause customers who are of little value to the company will reduce customer value for the company. therefore, a method is needed to find out the customer's value to the company. CLV method is a method that connects the value given by the company to the customer and the value that the customer gives to the company during the period of the relationship between the two. According to [8] Value of the customer can be seen by knowing the value generated at this time from all profits for the future of the company. While according to [6] the results obtained from the CLV model can be used to evaluate the future value of customers to companies with data mining techniques in detecting patterns and relationships in customer data history.

RFM (Recency, Frequency and Monetary) is one model that can measure CLV value, which is able to estimate the life value of customers [9],[10]. In addition, the advantage is that this model can also measure the value of customer profitability. [11]. The first time the RFM Model was developed by Hughes (1994), is one of the most common segmentation method that can identify customer values in companies with 3 variables: recency, communication and monetary [12]. Recency calculation shows the time since the last customer purchase transaction, frequency indicates the number of purchase transactions, while monetary shows the value of the purchase [13]. Then the RFM model is extended with one other variable: Length by Chang and Tasy (2004) [5] which is called the LRFM model (Length, Recency, Frequency and Monetary) [14]. This variable length is a variable that measures the length of a customer's first and last transaction at a time period. This is important to measure in strengthening the value of recency generated by customers based on the length of the last transaction with a certain time period. [5]. The LRFM model has the advantage of being more appropriate for analyzing customer life time. The disadvantage of the RFM model is that it does not consider customer loyalty, and this has been revised in the LRFM model. [15].

The main step in measuring the value of customer profitability is identifying customers by using the customer segmentation approach [9]. In conducting the segmentation process, existing customers are divided into several groups according to certain standards. [11]. Data mining is a science that can be used to do the segmentation process. [16] “Data mining is a series of processes to find relationships between items to form new knowledge patterns from large-sized data and processed with several techniques involving sciences including statistics, and mathematics.”. CLV is part of the Customer Relationship Model (CRM) activity. Data mining can extract information about customers that are important for effective strategy building (CRM). CRM activities that can be supported through the use of data mining include: (1) customer segmentation will produce customer segments based on customer behavior characteristics, (2) direct marketing campaigns to communicate directly with customers in increasing the number of new customers and purchasing products, and (3) market basket analysis to identify related products that are usually purchased together. [17].

The method used in customer segmentation is the clustering method. Clustering is also referred to as data segmentation in several applications because grouping large data into groups has similarities with the group.[18]. Clustering is the process of grouping physical and abstract datasets from objects into groups of objects that have similarities. [19], [20]. Clustering is a method that is widely used in various fields including customer segmentation, customer behavior, customer profitability, data collection of forest fires and so on. Various algorithms are used in clustering: K-Means, DBSCAN (Density Based Spatial Clustering of Application with Noise), SOM (Self Organizing Map) and others. The most commonly used method is the K-Means algorithm. [5].

However, K-Means clustering has a weakness in the accuracy of determining the number of clusters. [21], [5], [13], [12]. Many researchers have examined the validation method in determining the best number of clusters. One of the good methods for determining the cluster number is the Elbow Method. [22]. To determine the number of clusters using the Elbow method. This method is used in cluster analysis for interpretation and performance testing of the level of consistency of the right number of clusters by looking at the SSE value. [23]. At a certain point the graph will decrease dramatically with a curve called the elbow criterion. The value then becomes the best value of k or number of clusters. [23].

Therefore this study aims produce Customer Lifetime Value (CLV) with the K-Means clustering algorithm based on LRFM models (Length, Recency, Frequency and Monetary). The dataset used as an experiment is SME data sales of all operator pulses over a period of 1 January 2018 to 30 June 2018. Before the clustering process is carried out normalization and transformation are carried out to produce a dataset ready for the clustering process. Determining the best number of clusters is calculated by the elbow method. Then CLV results from the weighting of the LFRM data and the analysis and ranking process.

2. Related Works

2.1 Customer Lifetime value (CLV)

According to Cheng, defining customer value is a model based on calculating the distance between zeros and center clusters as a high value and refers to most customer loyalty inside. [21]. CLV is used in calculating the value of customer profitability. CLV is calculated after segmenting customers. The CLV value is calculated based on the CLV ranking specified for each segment. [8]. CLV calculation uses the equation i.

$$C^j = W_L C_L^j + W_R C_R^j + W_F C_F^j + W_M C_M^j \quad (1)$$

Where:

C^j = Customer CLV rating j

$C_L^j\ C_R^j\ C_F^j\ C_M^j$ = Normalization L, R, F, M from cluster j

$W_L\ W_R\ W_F\ W_M$ = result of weight L, R, F, M from AHP

2.2 Legth, Recency, Frequency, Monetery (LRFM)

According to [24] the LRFM model is a development of the RFM model developed by Arthur Hughes (1994). RFM is a model that is widely used in segmenting customer behavior. This RFM model consists of Recency, Frequency and Monetary. Next Chang and Tsay add one variable, namely Length. [5]. According to [24] , the definition of LRFM is as follows:

1) Length isvto measure relationship length between the company and the customer during the analysis period
2) Recency, which is the last date a transaction was made by a customer to the company during the analysis period.
3) Frequency is the number of transactions carried out by customers to companies during the period analyzed
4) Monetary, namely the amount of money that customers spend on companies during the analysis period .

The segmentation method with LRFM gives value according to customer value. [24]. The customer value will be segmented according to each segment in 16 clusters found in Figure 1. [14].

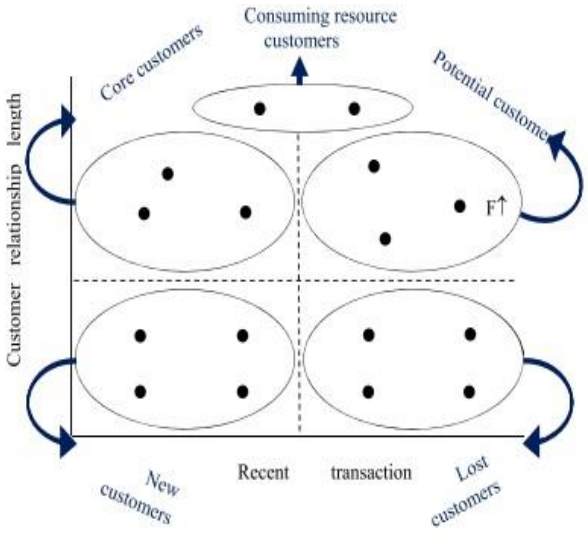

Fig. 1- Customer loyalty matrix basis (Chang & Tsay, 2004) in [14]

2.3 Clustering

K-Means is a method that is categorized into partition clustering method. [24], [25]. This condition is caused because this algorithm has simplicity and ease of use and users can determine the number of clusters themselves. [24]. There have been many roles of clustering in segmenting customers based on RFM or LRFM. In a research study of the literature review conducted by previous researchers, it appears that the role of K-means clustering techniques, C-means clustering, hierarchical clustering, and so on are usually used for the analysis of customer purchasing behavior grouping. Can be seen in Table 1. [14].

Table 1- Review of previous researcher [14]

Research	Indices	Clustering method
Hughes (1994)	RFM	-
Miglautsch (2000)	RFM	-
Shih and Liu (2003)	RFM	K-means clustering
Chang and Tsay (2004)	LRFM	Self-organizing maps (SOM)
Hu and Jing (2008)	RFM	K-means clustering
Bin, Peiji, and Dan (2008)	RFM	K-means clustering
Wu et al. (2009)	RFM	K-means clustering
Chang et al. (2010)	RFM	K-means clustering
Li et al. (2011)	LRFM	Two-Step clustering
Wei et al. (2012)	LRFM	SOM
Chen (2012)	RFM	C-means clustering
Kafashpoor and Alizadeh (2012)	RFM	Hierarchical Clustering

The steps in the K-means method are as follows: [24]:

[1].Determining the clusters number.
[2].Select the initial centroid randomly according to the number of clusters

[3].Calculate the distance of the data to the centroid using the euclidean distance formula.

$$D_{xy} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)$$

[4]. Update centroid by calculating the average value of values in each cluster.
[5]. Return to stage 3 if there is still data that moves clusters or changes in the value of the centroid

2.4 Elbow Method

The elbow method is used to determine the cluster's number from the dataset, by determining clusters so that total intra-cluster variation (total within variation of cluster or total within cluster sum of square) is minimized. This method is a visual method, starting with k = 2, and continuing to increase in each step with plus 1 at the value k. At the value of k = 3, if there is a drastic change that is inversely proportional to the previous value, then the value before the change is considered as the most appropriate number of clusters. [26]. The Elbow Method graph in Figure 2.

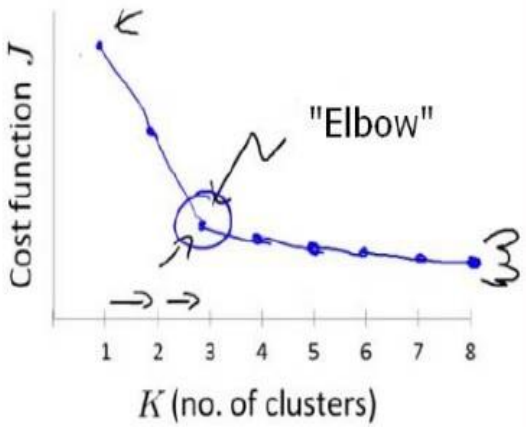

Fig. 2 - Elbow Graph to identification elbow point. [26]

3. Research Methodology

In this section the purpose model to CLV is described.

1). Select data according to the attributes L, R, F and M.
2). Normalizing the data due to the difference in data between L, R, F and M which is so far that it needs to be normalized by the min-max method.

$$V^1 = \frac{V - min}{max - min} (new\ max - new\ min) + new\ min \quad (3)$$

Where:

V = min max value
New max, new min = range min and max

3) Clustering with the best K-Means for k=1 to k=n
4).Determine the optimal number of clusters with the Elbow method
5) Choose the best cluster based on the elbow method.
6) Weighting the LRFM value using the AHP method.
7) Multiplying the LRFM weight value by an average the LRFM value and add it as shown in equation 8 below:

$$CLV = L * W_L + R * W_R + F * W_F + M * W_M \quad (4)$$

Where:

L,R,F,M = Average of customer value
L_w, R_w, F_w, M_w = weight of each value of LRFM

8) Ranking and analyzing cluster results with LRFM rank.

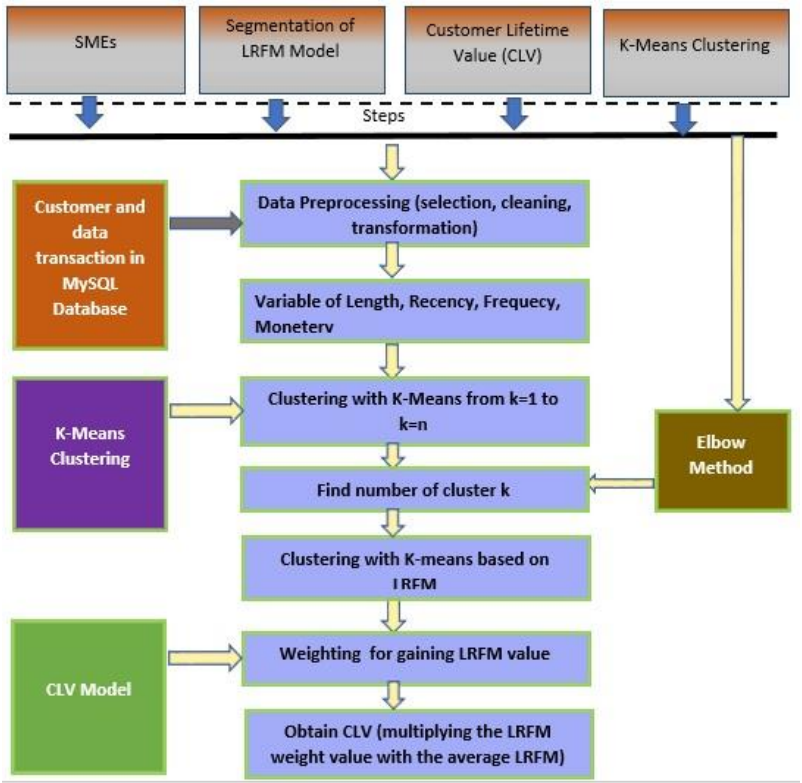

Fig. 3 - Segmentation model of Customer Lifetime Value based of K-Means clustering and LRFM

4. Experimental Result

4.1 Data Preprocessing Phase.

The study uses data from SME Customers all pulse servers operators AR-Pulsabiz Malang Indonesia. Total customers are 141 in the analysis period January 1, 2018 to June 30, 2018. The data will first be carried out in the form

of filtering query preprocessing process and transformation. The result of dataset after preprocessing described in Table 2 and normalization had been done and shown in Table 3.

Dataset of customer based on L,R,F,M before normalization described in Table 2.

Table 2 - Dataset before normalization

Id	Length	Recency	Frequency	Monetery
AR0001	179	179	184	3969775
AR0004	179	179	430	7175125
AR0008	179	179	5207	109052900
AR0010	176	176	656	13453050
AR0011	164	171	19	195450
AR0012	178	179	286	9924175
AR0016	164	173	38	850100
AR0018	178	179	48	980250
AR0024	179	179	1038	31893400
AR0028	179	179	295	7176275
..	..	..	..	..
AR2340	25	179	108	1271725

Dataset of customer based on L,R,F,M after normalization described in Table 3.

Table 3- Dataset after normalization

Id	Length	Recency	Frequency	Monetery
AR0001	1	1	0.035152	0.036351
AR0004	1	1	0.082405	0.065745
AR0008	1	1	1	1
AR0010	0.98324	0.98324	0.125816	0.123316
AR0011	0.916201	0.955307	0.003458	0.001739
AR0012	0.994413	1	0.054745	0.090955
AR0016	0.916201	0.96648	0.007107	0.007742
AR0018	0.994413	1	0.009028	0.008936
AR0024	1	1	0.199193	0.29242
AR0028	1	1	0.056473	0.065755
..	..	..	..	..
AR2340	0.139665	1	0.020553	0.011609

4.2 K-Means Clustering Phase.

In the process of counting with K-Means, 5 clusters have been formed, namely 1, 2, 3, 4, and 5. The following are the results of graphic processing and visualization of each cluster.

Table 4 - Cluster with k-1

Mean/ Centroid	***Ln***	***Rn***	***Fn***	***Mn***
AVERAGE	0.80	0.83	0.06	0.05
Respondents	**Number**	**%**	**SSE/Segment**	
Segment 1	100	100.0%	19.4	
TOTAL	100	100.0%	Average = 19.4	

Table 5 - Cluster with k=2

Mean/ Centroid	***Ln***	***Rn***	***Fn***	***Mn***
Segment 1	0.31	0.42	0.02	0.01
Segment 2	0.97	0.98	0.07	0.07
AVERAGE	0.80	0.83	0.06	0.05

Respondents	**Number**	**%**	**SSE/Segment**
Segment 1	26	26.0%	2.3
Segment 2	74	74.0%	2.8
TOTAL	100	100.0%	Average= 5.0

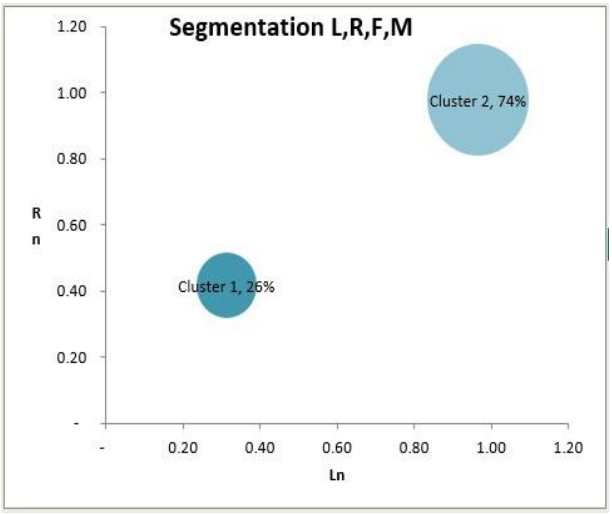

Fig. 4 - Visualization graph of segmentation with k=2

Table 6 - Cluster with k=3

Mean/ Centroid	***Ln***	***Rn***	***Fn***	***Mn***
Segment 1	0.27	0.38	0.02	0.01
Segment 2	0.95	0.97	0.06	0.05
Segment 3	1.00	1.00	1.00	1.00
AVERAGE	0.80	0.83	0.06	0.05

Respondents		
Number	**%**	**SSE/Segment**
Segment 1 :23	23.0%	1.7
Segment 2 :76	76.0%	1.6
Segment 3: 1	1.0%	0.0
TOTAL : 100	100.0%	Average = 3.3

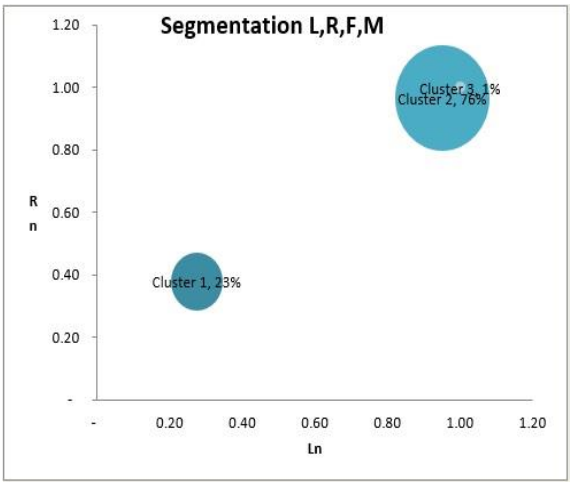

Fig. 5 - Visualization graph of segmentation with k=3

Table 7- Cluster with k=4

Mean/ Centroid	***Ln***	***Rn***	***Fn***	***Mn***
Segment 1	0.99	0.99	0.15	0.15
Segment 2	0.31	0.42	0.02	0.01
Segment 3	0.96	0.98	0.03	0.03
Segment 4	1.00	1.00	1.00	1.00
AVERAGE	0.80	0.83	0.06	0.05
Respondents	**Number**	**%**	**SSE/Segment**	
Segment 1	15	15.0%	0.3	
Segment 2	26	26.0%	2.3	
Segment 3	58	58.0%	0.4	
Segment 4	1	1.0%	0.0	
TOTAL	100	100.0%	Average = 2.9	

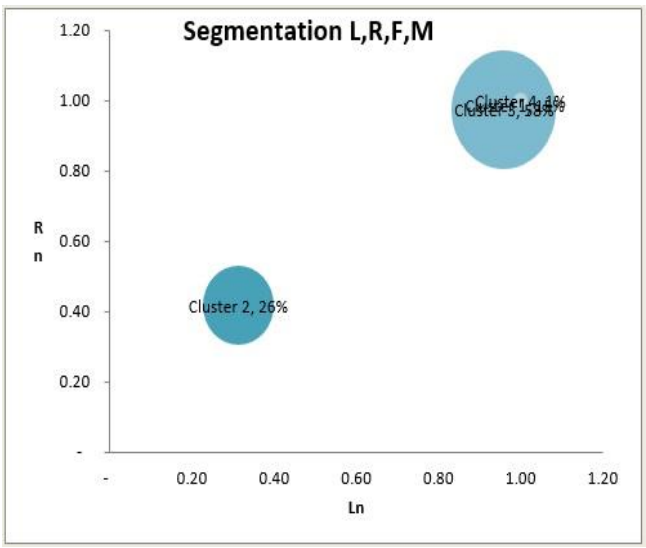

Fig. 6 - Visualization graph of segmentation with k=4

Table 8 - Cluster with k=5

Mean/ Centroid	***Ln***	***Rn***	***Fn***	***Mn***
Segment 1	1.00	1.00	1.00	1.00
Segment 2	0.99	0.99	0.13	0.13
Segment 3	0.98	0.99	0.03	0.03
Segment 4	0.26	0.33	0.02	0.01
Segment 5	0.73	0.84	0.01	0.01
AVERAGE	0.80	0.83	0.06	0.05

Respondents	**Number**	**%**	
Segment 1	1	1.0%	**SSE/Segment**
Segment 2	20	20.0%	0.0
Segment 3	42	42.0%	0.3
Segment 4	20	20.0%	0.1
Segment 5	17	17.0%	1.0
TOTAL	100	100.0%	1.0
			Average = 2.3

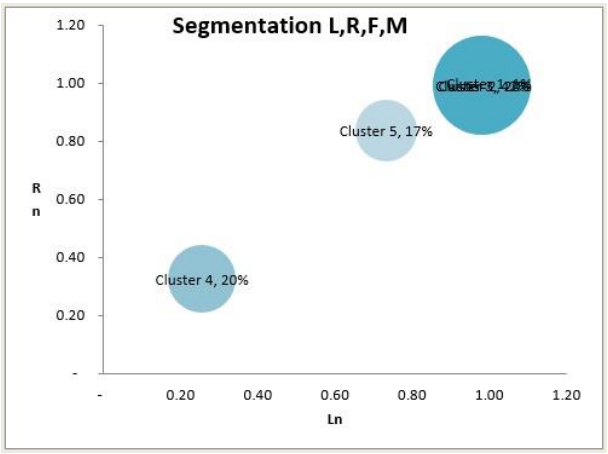

Fig. 7 - Visualization graph of segmentation with k=5

Figure 9 is a chart visualization of all customers and cluster centers, namely the main chart for 5 clusters.

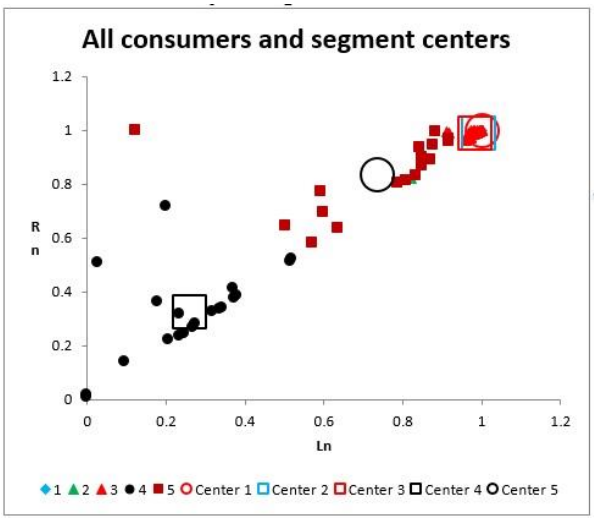

Fig. 8 - Central main chart for 5 cluster.

4.3 Select the best cluster with Elbow Method.

In this test the performance of each number of clusters is adjusted to the range of values in the Elbow method. The following are the results of the SSE performance test in Table 9. Figure 9 and Figure 10.

Table 9 - SSE result

Number of k	SSE	difference
1	19.4	
2	5.0	14.4
3	3.3	1.7
4	2.9	0.4
5	2.3	0.6

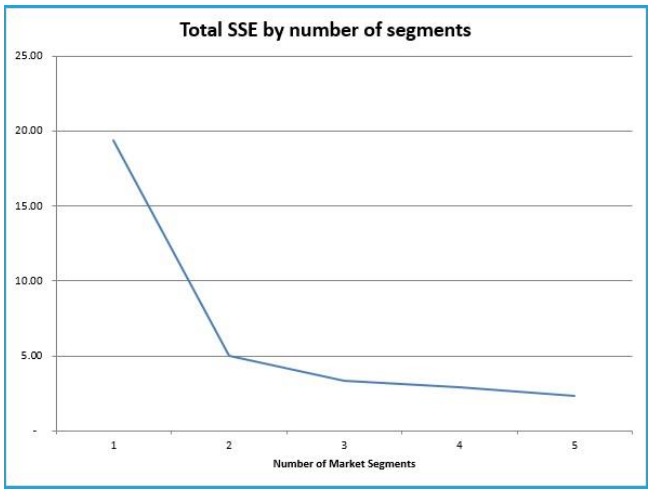

Fig. 9 - Elbow Graph Result

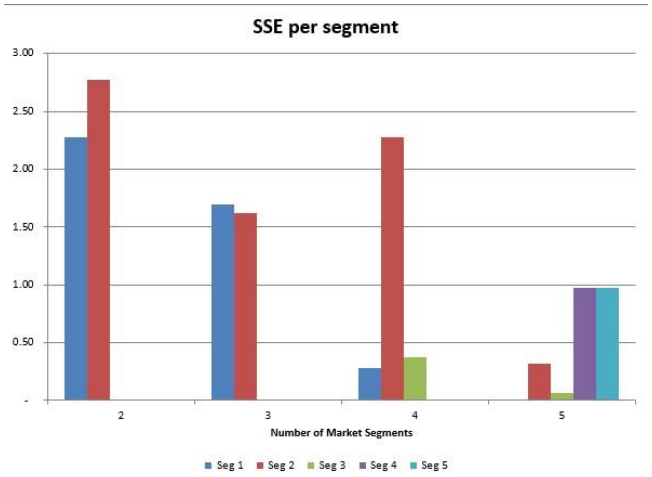

Fig.10 - Elbow Graph number of market segment.

Based on Table 9, the decline is seen in cluster-1 and cluster-2, which are marked with the most difference values with a value of 14.4, while at the next point there is a steady decline. Then the value of k used is 2.

4.4 Processing CLV Model

The generated cluster will be multiplied by the weight of the LRFM value using the AHP (Analytical Hierarchy Process) weighting method. In this case, the researcher used the same weight as the study [9] with the values of WL, WR, WF, WM, namely 0.238, 0.088, 0.326, and 0.348. This weight is used because each company has its own opinion regarding the priorities of each attribute L, R, F and M. This research customer [9] is a retail customer. This is the same as the customer studied in this study.

Table 10 - CLV value on LRFM Model (Before multiplied by weight result)

Mean/Centroid	Number Of customer	Ln	Rn	Fn	Mn
Segment 1	26	0.313	0.417	0.019	0.014
Segment 2	74	0.966	0.978	0.070	0.067
AVERAGE		0.796	0.832	0.056	0.054

$W_L = 0.238, W_R = 0.088, W_F = 0.326, W_M = 0.348$

Table 11- Ranking CLV on each cluster in the LRFM model.

Mean/ Centroid	Number of Customer	Ln*W_L	Rn*W_R	Fn*W_F	Mn*W_M	CLV	Rank
Segment 1	26	0.075	0.037	0.006	0.005	0.122	2
Segment 2	74	0.230	0.086	0.023	0.023	0.362	1
AVERAGE		0.152	0.061	0.014	0.014	0.242	

Based on Table 10 and Table 11 thenthe customer segment that has the highest CLV value is in cluster 2 with CLV value = 0.362. Based on Figure 1 from the research of Chang and Tsay (2004), this segment has a high loyalty value with the symbol LRFM L ↑ R ↑ F ↑ M ↑ which is a loyal customer (the best segment that has high customer loyalty value). Based on the LRFM symbol, the company can make a strategy to retain customers and acquire customers to become loyal customers with high profitability.

5. Conclusion and Future Work

This study produced 2 of the best clusters with elbow and Sum Square Error (SSE) methods. The formed cluster has been analyzed to find the CLV value. It is useful in determining the profitability value and customer loyalty with the LRFM model. Based on the CLV value, the highest ranking among the 2 clusters is at the second cluster with the CLV value being far the highest from the other cluster average of 0.362. The second ranking is in cluster-1 with the value of CLV is 0.122. Cluster with the highest CLV value namely cluster 2 has a customer segment in the form of high loyalty with high length values, low recency, high frequency and high Monetary. This means that customers in this segment often make purchases with a high amount of money spent on the company so that this segment is said to be a customer with high profitability. Therefore, customers in this segment need to be maintained by providing the best service so that they are not targeted by competitors. For further research, it can be focused on determining the number of clusters because by looking at the weaknesses of the clustering technique in determining the number of clusters still using one method, and the validity of a better number of clusters needs to be tested with several methods and compared.

Acknowledgement

Author would like to thanks to everyone that supports this research starting from University Teknikal Malaysia Melaka (UTeM), Malaysia, DRPM KEMENRISTEK DIKTI, Widyagama University of Malang, Indonesia, AR-PulsaBiz Reload Malang, Indonesia which is the one of SMEs in Indonesia, and anonymous reviewers who gave some suggestions and valuable critiques for improving this paper.

References

[1] J. Di Russo, "Product Recommender System for Small-Scale Niche Businesses using Association Rule Mining," 2017.

[2] Y. R. Suci, S. Tinggi, and I. Ekonomi, "Perkembangan UMKM (Usaha Mikro Kecil dan Menengah) di Indonesia," *J. Ilm. Fak. Ekon.*, vol. 6, no. 1, pp. 51–58, 2017.

[3] T. Hong and E. Kim, "Segmenting customers in online stores based on factors that affect the customer's intention to purchase," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2127–2131, 2012.
[4] S. Jamali and Y. Dorostkar Navaei, "A two-level Product Recommender for E-commerce Sites by Using Sequential Pattern Analysis," *Int. J. Integr. Eng.*, vol. 8, no. 1, pp. 15–24, 2016.
[5] S. Monalisa, "Klusterisasi Customer Lifetime Value dengan Model LRFM menggunakan Algoritma K-Means," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, p. 247, 2018.
[6] F. A. Buttle and S. Maklan, *Customer relationship management: concepts and technologies*, no. October. 2016.
[7] C. S. D. Prasetya, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 194, 2017.
[8] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Procedia Computer Estimating customer lifetime value based on RFM analysis of customer purchase behavior : case study," *Procedia Comput. Sci.*, vol. 3, pp. 57–63, 2011.
[9] Amin Parvaneh and M. J. T. Hossein Abbasimehr, "Integrating AHP and Data Mining for Effective Retailer Segmentation Based on Retailer Lifetime Value," *J. Optim. Ind. Eng.*, vol. 11, pp. 25–31, 2012.
[10] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, 2018.
[11] X. He and C. Li, "The Research and Application of Customer Segmentation on E-commerce Websites," in *2016 6th International Conference on Digital Home The*, 2016.
[12] R. Ait, A. Amine, B. Bouikhalene, and R. Lbibb, "Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model : The Case of Online Stores in Morocco," vol. 9, no. 8, pp. 1976–1986, 2015.
[13] C. D. Rumiarti and I. Budi, "Segmentasi Pelanggan Pada Customer Relationship Management di Perusahaan Ritel: Studi Kasus Pt Gramedia Asri Media," *J. Sist. Inf.*, vol. 1, no. 13, pp. 1–10, 2017.
[14] A. A. Zoeram and A. K. Mazidi, "A New Approach for Customer Clustering by Integrating the LRFM Model and Fuzzy Inference System," *Iran. J. Manag. Stud.*, vol. 11, no. 2, 2018.
[15] A. A. Zoeram and A. K. Mazidi, "A New Approach for Customer Clustering by Integrating the LRFM Model and Fuzzy Inference System.," *Iran. J. Manag. Stud.*, vol. 11, no. 2, pp. 351–378, 2018.
[16] F. Nurdiyansyah, S. Arifin, and F. Marisa, "Penerapan Clustering Algorithm Untuk Mendukung Promosi Server Pulsa Reload," vol. 3, no. 2, pp. 73–78, 2018.
[17] G. S. L. Michael J. A. Berry, *Data Mining Techniques*, 2nd ed. Indianapolis, Indiana: Wiley Publishing, Inc, 2004.
[18] J. P. Jiawei Han, Micheline Kamber, *Data Mining – Concepts & Techniques*, Third. Waltham: Elsevier Inc., 2012.
[19] P. A. S. A. U. H. Takci, "Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis," *Kybernetes*, vol. 45, no. 7, p. , 2016.
[20] M. Z. Saringat, A. Mustapha, and R. Andeswari, "Comparative Analysis of Mice Protein Expression: Clustering and Classification Approach," *Int. J. Integr. Eng.*, vol. 10, no. 6, pp. 26–30, 2018.
[21] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty," *Expert Syst. Appl.*, vol. 37, no. 7, pp. 5259–5264, 2010.
[22] A. D. Savitri, F. A. Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, pp. 2957–2966, 2018.
[23] B. Purnima and K. Arvind, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014.
[24] D. Kandeil, A. Saad, and S. M. Youssef, "A Two-phase Clustering Analysis for B2B Customer Segmentation," in *2014 International Conference on Intelligent Networking and Collaborative Systems*, 2014.
[25] R. Rustiyan and M. Mustakim, "Penerapan Algoritma Fuzzy C Means untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, p. 171, 2018.
[26] P. R. M. Kodinariya, Trupti M, "Review on determining number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 90–95, 2013.