The Analyze of Relationship between Revenue and Customer Payment Methods in Small Medium Enterprise Based on Clustering K-Means

The Analyze of Relationship between Revenue and Customer Payment Methods in Small Medium Enterprise Based on Clustering K-Means

Fitri Marisa[1,2,5], Sharifah Sakinah Syed Ahmad[1], Zeratul Izzah Mohd Yusof[1], Tubagus Mohammad Akhriza[3], Wiwin Purnomowati[2], Rakesh Kumar Pandey[4]
[1] Faculty of Information and Communication Technoloy - Universiti Teknikal Malaysia Melaka, Jl. Hang tuah jaya 76100 durian tunggal Melaka,Malaysia.
[2] Informatics Engineering Department - Widyagama University of Malang, Jl. Borobudur no 35 Malang Jawa Timur – Indonesia
[3]Department of Infermatics Engineering - Pradnya Paramita School of Informatics and Management Malang, Indonesia
[4] Kirori Mal College – University of Delhi – North campus Delhi 110007- India
[5] Research fellow of Department of Science Technology (DST) under AIRTF Program -RTF/2018/000033 India

E-mail: fitrimarisa@widyagama.ac.id

Abstract. Business capital and revenue are not only the decisive of the health of SMEs but also they must be balanced. In general, customers find their benefit from the flexible payment methods while on the other hand the SMEs should get their benefit too. So that, it needs to be studied whether it is necessary for SMEs to get their profit in accordance to this situation. One of the methods that suitable to be applied is by applying customer groupings based on revenue and payment namely the K-means clustering method since it can raise several groups that have not been known before. This information is useful for SMEs to be utilized based on their needs. Data in this study were gathered from customer attributes, number of transactions, and payment methods. The number of centroids was 3. The grouping results were stopped at the 5th iteration. The finding showed that the ratio value of the 4th iteration and the 5th iteration having the same ratio value, 0.07393. From the results of the iterations can be found; first, based on the customers' number, the groups can be classified into three C1(18%), C2 (45%), C3 (36%). Second, based on the average number of transactions, post-paid payments was in the first rank (12.7 / week). From the results, it can be analyzed that this situation is burdensome for SMEs because the more the number of transactions, the more investment must be prepared for accounts receivable.

Keyword: *Clustering, Transaction, Payment-method, Small Medium Enterprise*

1. Introduction

SME is an asset for Indonesia, and is proven to survive the economic crisis. [1]. Therefore, it is important about how to help SMEs to develop themselves and solve their problems.[2]. The customer is an asset for Small Medium Enterprise (SME) [3],[4], where as an effort is the company to provide convenience for customers. Payment flexibility is one of the components needed by the customer, which can be an attraction for customers towards the company. However, the flexibility of payment for customers can

have consequences for the company by providing more capital. If the consequences are balanced by the amount of profit the consequences are balanced by the amount of profit the company receives, then this is not a problem or vice versa. Of course, it is necessary to do a study of how the relationship of the number of customer transactions with the customer payment system, therefore companies can make more informed decisions for the continuity and development of their business.
SMEs can use their transaction database to mine and analyze these needs.[5]. Most companies have realized that customer databases are very important resources that can be used to analyze customer characteristics to form appropriate marketing strategies and to regulate them.[6]. Customer segmentation is one of the activities commonly used by utilizing these data. In general, customer segmentation is defined as the process of grouping customers by being divided into several groups based on the characteristics and behavior of their transactions. [7]. Many studies have been conducted on customer segmentation. each study has a different purpose. One of study used K-Means find important ranked stock. [8], and other study to find customer segmentation user CRM and K-Means. [9]. This study has analyzed based on RFM (Recency, Frequency, Monetary) to find the potential customer. [10].

One approach is to extract data using the clustering method in data mining. Data mining contributes greatly to extracting hidden knowledge and information contained in the data used by researchers.[11]. The process of data mining in SME data is done by clustering using the K-means algorithm for customer segmentation. [12]. In research [13] has combined clustering methods with assembled algorithms to segment customer.

2. CLUSTERING METHODOLOGY IN DATA MINING

Knowledge discovery in Database (KDD) is defined as potential, implicit and unknown information extraction from a set of data [14], [15]. Datamining is in the KDD framework which is part of the component. K-means algorithm is an algorithm for grouping data that only applies to numerical data, so that if the data is not numeric normalization must be done.[16]. K-Means evective to reduce the euclidean distance from each object. [17].

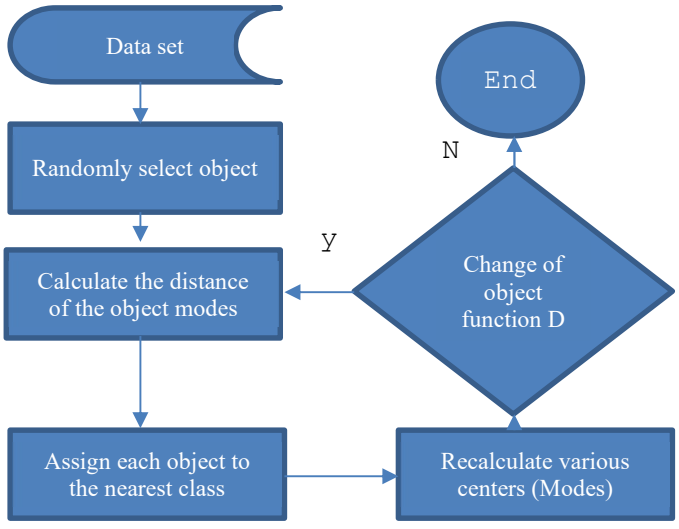

Figure 1: K-Means Algorithm Step.[17].

Several studies have been conducted among others: This study used the clustering method to cluster customer segmentation with RFM model.[18]. Other research tried to combine clustering method with decision tree for segmentation of customer.[19][20]. Development of clustering with multiple attributes

has also been done in marketing segmentation based on greedy heuristic. [21]. Segmentation of customer to analys of marketing use K-Means has done in [22] research. Other research, develop the clustering of customer segmentation to analys customer intention. [23][24].

3. Research Methodology

The purpose model of clustering of customer segmentation with K-Means can describe in the Figure 2.

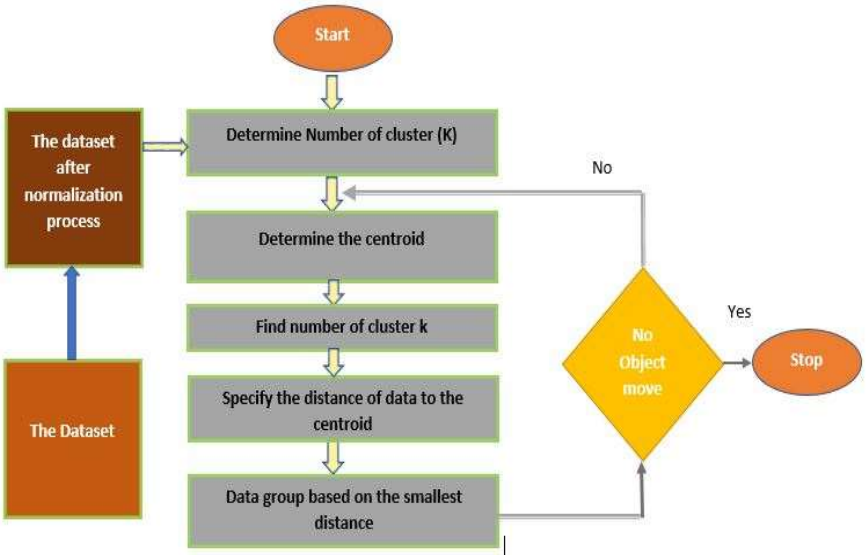

Figure 2: Methodology of customer segmentation

Each step of methodology in Figure 2 can be describe:

1. Determine the number of cluster by determine K value
2. Determine the centroid randomly
3. Find number of cluster K
4. Specify the distance of data to the controid

$$d_{ij} = \sqrt{\sum_{k-1}^{p}\{X_{ik} - X_{jk}\}}^{2} \qquad (1)$$

d_{ij} = Distance of object between object *i* and *j*
P = dimension of data
X_{ik} = The coordinates of object *i* in dimension *k*
X_{jk} = The coordinates of object *j* in dimension *k*

BCV (Between Cluster Variation) = eucledien distance from *m* to *mj*

$$BCV = d(mi,mj)$$

WCV (Within Cluster Variation) = the smallest distance between data and centroid.

$$WCV = \sum(\textit{the smallest distance between data and centroid})^2$$

5. Data group based on the smallest distance

4. Experiment and Result

The dataset is extracted from data from one of the SMEs engaged in the sale of electric pulses. The first step is to do a proprocessing dataset. The dataset (Table-1) is carried out by the normalization process by making the category of payment system data an integer value.

Table 1: Data set

Customer	Average of transaction/ week	Payment system
King	96	Auto-transfer
Al Hikam	352	Post-Date
Yuyun	34	Post-Date
Sigit	31	Non-Auto-transfer
Naga	7	Auto-transfer
Prata	34	Post-Date
Weni	28	Combination
Nurul	7	Combination
MD	15	Post-Date
Dince	27	Combination
ChandCell	10	Post-Date

There are 4 categories of payment method: Auto-transfer (100), Non-auto-transfer (75), Combination (50), and post-date (25). (Table-2).

Table 2: Dataset after Normalization

Customer	Average of transaction/week	Payment system
King	13.71	100
Al Hikam	50.29	25
Yuyun	4.86	25
Sigit	4.43	75
Naga	1.00	100
Prata	4.86	25
Weni	4.00	50
Nurul	1.00	50
MD	2.14	25
Dince	3.86	50
ChandCell	1.43	25

Section 1.01 Iteration-1

In this study, it has determined the number of clusters (K = 3). Iteration-1 by determining the number of centroids at random. The centroid are M1 {13.72,100}, M2 {50.29,25}, M3 {4.43,75} from table-2. The distance between centroids is calculated and then the shortest distance between the centers is selected as a determinant of the cluster group of data. (Table-3).

Table 3: The distance of data for iteration-1

Customer	C1	C2	C3	Distance^	Cluster
King	0	83.44	26.67	0	C1
Al Hikam	83.44	0	67.84	0	C2
Yuyun	75.52	45.42	50.00	2063.755	C2
Sigit	26.67	67.84	0	0	C3
Naga	12.71	89.74	25.23	161.6531	C1
Prata	75.52	45.43	50.00	2063.755	C2
Weni	50.93	52.61	25.00	625.1837	C3
Nurul	51.59	55.267	25.23	636.7551	C3
MD	75.89	48.14	50.05	2317.735	C2
Dince	50.96	52.73	25.00	625.3265	C3
ChandCell	75.99	48.86	50.09	2387.02	C2

The final step is to find the ratio that is the result of between BCV (Between Cluster Value) and WCV (Within Cluster Value) in Table-3. The BCV value is 177.9547, and the WCV value is 5551.102. Therefore, ratio of Table-3 is 0.032058.

Section 1.02 Iteration-2

In the iteration-2 until the iteration-n has a different way of determining the cluster than iteration-1. In this step, the cluster is determined by calculating the average of the groups formed in the previous iteration, in Table-4.

Table 4: Table for determine centroid in iteration-2

	cluster 1	
Customer	Average of transaction/ week	Payment system
King	13.71	100
Naga	1.00	100
mean	7.36	100
	cluster 2	
Al Hikam	50.29	25
Yuyun	4.86	25
Prata	4.86	25
MD	2.14	100
ChandCell	1.43	25
Mean	12.71	40
	cluster 3	
Sigit	4.43	75
Weni	4.00	50
Nurul	1.00	50
Dince	3.86	50
mean	3.32	56.25

The centroid determined from table-4 are M1 (7.36, 100), M2(12.71, 40), M3(3.32, 56.25). The next step is to determine the centroid distance as in the iteration-1. It continues until the n-iteration.

Table 4b describe the result of data distance. In this iteration-2, it has produced changes in cluster groups from the previous iteration. This situation continues to be repeated until there is no change in the cluster group formed.

Table 4a : The distance of data for iteration-2

Customer	C1	C2	C3	Distance	Cluster
King	6.357	60.01	44.97	6.36	C1
Al Hikam	86.417	40.46	56.41	40.46	C2
Yuyun	75.042	16.93	31.29	16.93	C2
Sigit	25.171	35.97	18.78	18.78	C3
Naga	6.357	61.13	43.81	6.36	C1
Prata	75.042	16.93	31.29	16.93	C2
Weni	50.113	13.26	6.29	6.29	C3
Nurul	50.403	15.40	6.67	6.67	C3
MD	75.181	18.35	31.27	18.35	C2
Dince	50.122	13.36	6.27	6.27	C3
ChandCe	75.234	18.77	31.31	18.77	C2

Table-5 has presented the BCV, WCV, and ratio values of all iterations.

Table-5: List of iteration

Iteration	**BCV**	**WCV**	**Ratio**
Iteration-1	177.9547	5551.102	0.032058
Iteration-2	122.9438	2727.671	0.045073
Iteration-3	123.7195	2924.418	0.042306
Iteration-4	151.7579	2052.671	0.073932
Iteration-5	151.7579	2052.671	0.073932

The result of the iteration-4 ratio is the same as the iteration-5. It can be concluded that the iteration process has been completed and the cluster result was produced by the last iteration.
The result of customer segmentation based on clustering method as follows in Table 6:

Table 6: Customer Cluster

Customer	Aveage of transaction/week	Payment Method	Cluster
King	13.71428571	100	C1
Al Hikam	50.28571429	25	C2
Yuyun	4.857142857	25	C2
Sigit	4.428571429	75	C3
Naga	1	100	C1
Prata	4.857142857	25	C2
Weni	4	50	C3
Nurul	1	50	C3
MD	2.142857143	25	C2
Dince	3.857142857	50	C3
ChandCell	1.428571429	25	C2

It can be analysed how payment method give the impact for transaction average. This study has compared the average of transactions in each cluster which are related to the payment method, can be seen in Table 7 and Graph-1.

Table 7: Comparison of cluster in payment method and average of transaction

Cluster of payment method	Average of Transaction	Percentage / %
C1 (Auto-Transfer)	7.357142	18
C2 (Post-Date)	12.71428	45
C3 (Non-auto-transfer and Combination)	3.321428	36

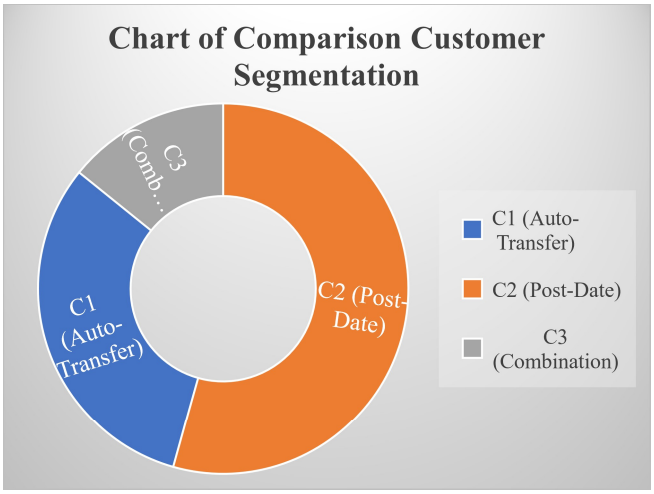

Figure 3 : Chart of Comparison of cluster in payment methods and average transaction

The grouping results have produced 3 clusters with the highest average transaction C2 use post-date payment (12.714), followed by C1 use Auto-Transfer payment (7.357), and C3 use Non-Auto-transfer and Combination payment (3.321). From the results of the grouping it appears that the biggest contributor to SME transactions is the customer with the post-date payment system. This data can be used as an important finding for SMEs to make decisions in the development and evaluation of their companies. From these results it can be analyzed that this situation is quite burdensome to SMEs because the more the number of transactions they will spend on the accounts receivable. Therefore, the clustering method is suitable for extracting data from companies to explore opportunities that might be used by companies including small and medium companies for evaluation and development.

4. Conclusion and Future Research

The clustering approach has been able to explore SME data to analyze possibilities that can be used as judgment to make decisions in the development and evaluation of the company. From the results of the iterations can be found; first, based on the customers' number, the groups can be classified into three C1(18%) is auto-transfer payment, C2 (45%) is post-date payment, C3 (36%) is non-auto-transfer and combination payment. Three (3) clusters formed resulted in the finding that the most transaction contributors were customers with a post-date payment system. Based on the average number of transactions, post-date payments was in the first rank (12.7 / week). This finding can provide advice for SMEs that many transactions by customers should also be balanced with the availability of capital. The smallest number of transactions is the customer with an auto-transfer system means that customers have

a tendency to be less interested in the payment system at the beginning, in other hand, they rely more on capital from SMEs. This can be a concern for SME managers to react to it. Determination of the number of clusters in this study is still done manually. In future work, it is necessary to think about the validity of the right number of clusters so that the results of grouping will be more accurate.

*5. **Acknowledgment***
Author would like to thanks to everyone that supports this research starting from University Teknikal Malaysia Melaka (UTeM), Malaysia, DRPM DIKTI, Department Science and Technology (DST) under AIRFT India (RTF/2018/000033), Widyagama University of Malang, Indonesia, AR-PulsaBiz Reload Malang, Indonesia which is the one of SMEs in Indonesia, and for collegiate who gave some suggestions and valuable critiques for improving this paper.

6. Reference

[1] A. Hadjimanolis, "Barriers to innovation for SMEs in a small less developed country (Cyprus)," *Technovation*, vol. 19, no. 9, pp. 561–570, 1999, doi: 10.1016/S0166-4972(99)00034-6.

[2] I. Janita and W. K. Chong, "Barriers of B2B e-business adoption in Indonesian SMEs: A literature analysis," *Procedia Comput. Sci.*, vol. 17, pp. 571–578, 2013, doi: 10.1016/j.procs.2013.05.073.

[3] 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, doi: 10.25126/jtiik.201852690.

[4] F. Marisa *et al.*, "Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model," *Int. J. Integr. Eng.*, vol. 11, no. 3, pp. 169–180, 2019.

[5] 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, doi: 10.25126/jtiik.201743392.

[6] 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, doi: 10.22059/ijms.2018.242528.672839.

[7] D. Kandeil, A. Saad, and S. M. Youssef, "A Two-phase Clustering Analysis for B2B Customer Segmentation," 2014, doi: 10.1109/INCoS.2014.49.

[8] H. W. Shin and S. Y. Sohn, "Segmentation of stock trading customers according to potential value," *Expert Syst. Appl.*, vol. 27, pp. 27–33, 2004, doi: 10.1016/j.eswa.2003.12.002.

[9] W. Qadadeh and S. Abdallah, "Customers Segmentation in the Insurance Company (TIC) Dataset," *Procedia Comput. Sci.*, vol. 144, pp. 277–290, 2018, doi: 10.1016/j.procs.2018.10.529.

[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, doi: 10.1016/j.jksuci.2018.09.004.

[11] 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, doi: 10.1016/j.eswa.2009.12.070.

[12] 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.

[13] R. J. Kuo, C. H. Mei, F. E. Zulvia, and C. Y. Tsai, "An application of a metaheuristic algorithm-based clustering ensemble method to APP customer segmentation,"

Neurocomputing, vol. 205, pp. 116–129, 2016, doi: 10.1016/j.neucom.2016.04.017.

[14] A. I. Warnilah, "Analisis Algoritma K-Means Clustering Untuk Pemetaan Prestasi Siswa," *Indones. J. Comput. Inf. Technol.*, vol. 1, no. 1, pp. 83–92, 2016.

[15] F. Marisa, "Educational Data Mining (Konsep Dan Penerapan)," *Teknol. Inf.*, vol. 4, no. 2, pp. 90–97, 2013.

[16] K. Yang and R. Miao, "Research on Improvement of Text Processing and Clustering Algorithms in Public Opinion Early Warning System," *2018 5th Int. Conf. Syst. Informatics, ICSAI 2018*, no. Icsai, pp. 333–337, 2019, doi: 10.1109/ICSAI.2018.8599424.

[17] H. Ismkhan, "I-k-means−+: An iterative clustering algorithm based on an enhanced version of the k-means," *Pattern Recognit.*, vol. 79, pp. 402–413, 2018, doi: 10.1016/j.patcog.2018.02.015.

[18] 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, doi: 10.1016/j.jksuci.2018.09.004.

[19] K. Khalili-Damghani, F. Abdi, and S. Abolmakarem, "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries," *Appl. Soft Comput. J.*, vol. 73, pp. 816–828, 2018, doi: 10.1016/j.asoc.2018.09.001.

[20] A. V Nair, A. Rahman, M. Kasim, and M. Z. Salleh, "Annals of Mathematical Modeling , 2 (2), 2020 , 43-53 A review on the fluid structure interaction of circular plates using numerical methods," vol. 2, no. 2, pp. 43–53, 2020.

[21] D. L. Huerta-Muñoz, R. Z. Ríos-Mercado, and R. Ruiz, "An iterated greedy heuristic for a market segmentation problem with multiple attributes," *Eur. J. Oper. Res.*, vol. 261, no. 1, pp. 75–87, 2017, doi: 10.1016/j.ejor.2017.02.013.

[22] F. Nurdiyansyah, S. Arifin, and F. Marisa, "Clustering algorithm untuk pengelompokan pelanggan dalam bidang usaha server reload," pp. 1043–1047, 2018.

[23] 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, doi: 10.1016/j.eswa.2011.07.114.

[24] N. Raihan, M. Asimoni, N. F. Mohammad, A. Rahman, and M. Kasim, "Annals of Mathematical Modeling , 1 (2), 2019 , 81-88 Unsteady MHD free convective flow past a vertical plate : An automated solution approach," vol. 1, no. 2, pp. 81–88, 2019.