Three-Layer Data Clustering Model for Multi-View Customer Segmentation using K-Means

Afgan Fazri Handoko, Antoni Wibowo

Abstract: Customer Relationship Management (CRM) system is one of the methods to increase customer satisfaction with the services provided by the company. The data in a CRM system sometimes have not been utilized properly to find specific information about customer needs. The data mining process can help companies to segment and retrieve useful information about customers. The segmentation of customers can be categorized into groups based on the RFM (Recency, Frequency, and Monetary) values of the customers. Several studies have used the RFM model as a basis for customer segmentation. However, the methods proposed in previous studies are very specific to certain industries and the range of RFM scores used is also very subjective. Also, as the business grows there are challenges with RFM score measurement. RFM score measurement needs frequent adjustments in which this adjustment is not easy using the existing methods. Therefore, this study proposed a novel method to overcome the limitation of the existing methods using combined K-Means and Davies-Bouldin Index (DBI) to find the appropriate range of RFM scores. Based on our study in a telecommunication industry the proposed method simplify the measurement of the RMF scores as the data grows. This research also provided the appropriate RFM score range through the K-Means approach based on the optimal K value of the K-Means algorithm. Our proposed method could be implemented in other industries since it only depends on the values of RFM from the correspond data for each customer.

***Keywords**: Clustering, Data mining, RFM Analysis, Segmentation.*

I. INTRODUCTION

Customers are the most valuable assets of a company, to retain customers the company is required to provide satisfactory services to customers. Many efforts made by the company to ensure customer satisfaction, one of which is to adopt a CRM system to maintain and use critical information regarding the needs, expectations, and choices of customers to make them satisfied and loyal [1]. With the customer transaction data contained in the CRM system, we can process the data using data mining techniques to obtain detailed knowledge and information for each customer [2]. One method that can be used to understand customers in large-scale CRM database is by segmenting each customer based on data similarity and place each customer in their respective segments [3].

The data mining technique that can be used for customer segmentation is clustering [4]. Clustering is the task of grouping unlabeled data points into multiple numbers of groups based on the similarity of the data points [5]. Clustering techniques using various algorithms such as c-means, K-Means, k-medoids have been proven capable of segmenting customers appropriately [6], [7], [8], [9].

Various approaches can be taken to segment customers. For example, the customer segmentation model can be based on ARPU (Average Revenue Per User), MOU (Minutes of Usage), DOU (Data of Usage), demographic and behavioral data of telecom customers [10]. Also, CLV (Customer Lifetime Value) can be used to segment customers [11]. And there also research that uses the RFM model as an approach to segment and explores specific information for each customer [12].

RFM is a model used for analyzing customer value. RFM stands for the three data layer: (1) Recency - How recently did the customer purchase; (2) Frequency - How often do the customer purchase; (3) Monetary Value - How much do the customer spend [13]. Through these three data layers, multi-view customer segmentation can be obtained. Although there have been many studies using the RFM method, there is no standard that can determine the RFM score range.

This research was conducted to create a customer segmentation model with three-layer data clustering based on the RFM model to produce comprehensive information about customers through multi-view customer segmentation. In addition, this research will show how to determine the range of scores on RFM using clustering techniques. The contributions of this paper are as follows: This research presents a method for generating the RFM values based on service subscription transaction data in the telecommunication industry in which we provide a novel method to define the appropriate RFM score range through K-Means approach based on the optimal K value of K-Means algorithm. Our method also simplifies the RFM scoring process when we add new data in the RFM process.

The remainder of this paper proceeds as follows. Section 2 discusses the previous approach in the RFM customer segmentation model. Section 3 describes the research methodology. Section 4 shows the experimental results. Section 5 presents the findings of this research and suggestion for future research.

Revised Manuscript Received on February 27, 2020.

Afgan Fazri Handoko*, Computer Science Department, BINUS Graduate Program–Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia. Email: afgan.handoko@binus.ac.id

Antoni Wibowo, Computer Science Department, BINUS Graduate Program–Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia. Email: antoni.wibowo@binus.edu

II. RELATED WORKS

To generate an RFM score, RFM values have often been used as input variables to segment customers. RFM values can be used as an input variable to produce a single cluster label [14], [15], [16]. For example, auto insurance customers can be segmented using K-Means and RFM values as the input parameter produced by the fuzzy analytic network process (FANP) [15].

Table-I: Related Works

Author	Data	Approach	RFM Scoring Method
[14]	Customers data from a telecommunication company	Segment RFM values using K-Means to produce single cluster label	RFM score produced in segment level using a mathematical formula
[15]	Customers data from an auto insurance company	Producing each RFM values using fuzzy ANP then applying K-Means to produce single cluster label	None
[16]	Prospective customers data from automobile retailer company	Using PSO algorithm to segment customers based on their RFM score to produce a single cluster label	RFM score produced using a logical formula to set the score range from 1 to 5
[17]	Customer transaction data from office stationery company	Using the K-Means algorithm to segment customers based on their RFM score & RFM Values to produce two cluster labels	RFM score produced using a logical formula to set the score range from 1 to 5
[18]	Customers transaction data from IT Company	Segment customer using FCM clustering algorithm then calculates RFM score for each cluster using fuzzy AHP	RFM score produced using a mathematical formula to set the score range from 0 to 1
[19]	Purchasing data from a branch general store	Segment customer using Firefly Algorithm (FA) then calculate RFM score in each cluster to find the most profitable segment	RFM score produced in segment level using a mathematical formula
[20]	Telecom customer service usage history data from a telecommunication company	Statistic-based approach to value latent users via time series interval usage as RFM values	RFM score produced using a mathematical formula to set the score range from 1 to 8 based on a statistical analysis
[21]	Customer transaction data from an e-commerce retail store	Using K-Means to segment each customer based on RFM score. The prediction of churn is then done using logistic regression, SVM and SGD classifier.	None

Customer segmentation can also be done using Particle Swarm Optimization (PSO) algorithm to segment customers based on each customer's RFM score [16]. Another approach to segment the customers can also be done by providing RFM score and RFM values as the input parameters so that it becomes 2 cluster labels [17]. An RFM score can also be produced at the segment level to find out which segment provides the greatest profit for the company [18], [19].

There several ways to produce RFM values and RFM scores, for example, using usage data from telecommunications services that are used by the customer can be converted to RFM values through a statistical approach [20]. The results from the RFM model can also be used as a parameter to predict customer churn [21].

As we can see in the previous research that has been mentioned, several authors have used the RFM model as a basis for customer segmentation. In previous research, there was no attempt to use clustering techniques to determine the range of RFM scores. Some previous research uses logical or mathematical formulas to determine the range of RFM score, with this approach the formulas need to be adjusted frequently as the data grows. Also, the methods proposed in previous studies are very specific to certain industries and the range of RFM scores used in previous studies is also very subjective [14], [15], [18], [19].

This research will focus to resolve the shortcomings that existed in previous research by using clustering techniques especially the K-Means algorithm to produce RFM scores based on the RFM values of each customer and finding the optimal K values for each of these three-layer data.

III. RESEARCH METHODOLOGY

This study aims to create a multi-view customer segmentation model based on the RFM model using Recency, Frequency and Monetary values of each customer. This research starts from attribute selection, data pre-processing, RFM values generation, RFM scoring, and evaluation. The research process is carried out as shown in Fig. 1.

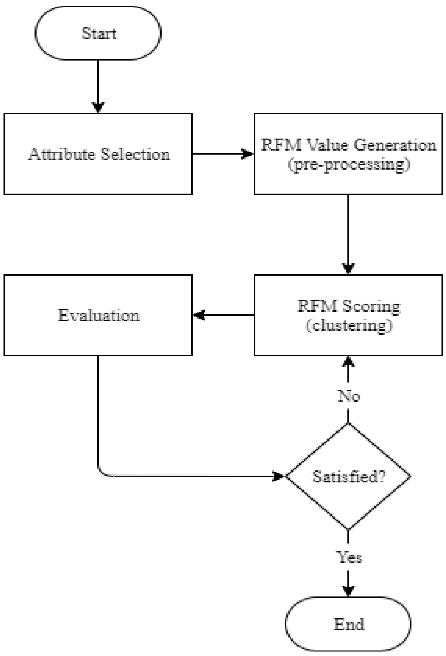

Fig. 1.Research methodology

A. Attribute Selection

Usually, the database contains a hundred attributes; the database may contain irrelevant attributes for the data mining process or data mining purpose. In this step, we choose the attributes needed to produce RFM values. In the Recency layer, we take attributes that indicate the date of the transaction. In the Frequency layer, we take the customer id attribute which will be counted according to the number of transaction records of each customer. Finally, in the Monetary layer, we take the price value of the transaction.

Table-II: Attribute Selection

Data Layer	Attribute	Description
Recency	Transaction Date	An attribute that indicates the first time a customer makes a transaction and the date on which the customer starts subscribing for a service at the company
Frequency	Customer Id	A unique indicator that identifies each customer
Monetary	Transaction Price	An attribute that contains the number of costs that customers must pay to the company each month

B. RFM Values Generation (pre-processing)

After selecting the attributes, on this step, we perform attribute transformation, generation, and normalization. Because each data layer has different attribute functions and different data types, we do the pre-processing part at each data layer separately. The data we use in this research come from the CRM system of a telecommunication company. The transaction is based on subscribed services and will be billed to customers every month.

For generating Recency value, we need to generate a standard period (billing period) which we can get through the transaction date attribute. Then the transaction date will be compared with the current date to find the duration of the current subscription period and converted to a monthly format.

For the second layer, which is Frequently value. We group records based on customer id attributes and add up the total records for each customer. By doing this we can get the number of transactions each customer has made with the company.

For the Monetary layer, we multiply the price of each transaction by the billing cycle duration and then add it up for each customer to get the total value of all transaction that has been paid by the customer to the company. The Monetary value of each transaction is calculated as (1) and monetary value for each customer calculated as (2).

$$M_i = P_i \times C_i \qquad (1)$$

$$M = M_{i^1} + M_{i^2} + \ldots + M_{i^n} \qquad (2)$$

Where i is transaction record, M is monetary value P is price value for a transaction record and C is the number of total cycle period the customer subscribing to the service.

After finding RFM values then we normalize the data separately on each data layer before processing the data using the K-Means algorithm.

C. RFM Scoring (clustering)

The core idea of K-Means is to update each cluster center represented by the center of data points by iterative computation until some criteria for convergence is met.

In K-Means the number of clusters K is an input parameter, for each data vector this algorithm calculates the distance between data vector and each cluster centroid as in (3) where Zp is p^{th} data point MJ is the centroid of j^{th} cluster. Then the centroid is recalculated each time respectively after the addition of data point in cluster j. It is calculated as in (4) where Nj is the number of the data point in cluster j.

$$D(Zp, Mj) = \sqrt{(\sum(Zp, Ky - Mj, Ky)} \qquad (3)$$

$$Mj = 1/Nj \sum Zp, \nabla Zp \in Cj \qquad (4)$$

Then there is the K-Means algorithm with input K_y as the number of clusters and D_y as a data set containing n object. The output is a set of K_y clusters.

1) Input the data set and value of K_y.
2) If K_y== 1 then Exit.
3) Else
4) Choose k objects from D randomly as the initial cluster centers.
5) For every data point in the cluster, j reissue and define every object into the cluster where the object matches, based on the object's mean value in the cluster.
6) Update cluster means; after that for each cluster calculate the object's mean value.
7) Repeat from step 4 until no data point was assigned otherwise stop.

The satisfying criteria can be either several iterations or change of position of the centroid in consecutive iterations [22]. After finding the cluster label on each data layer then we apply our proposed method of RFM scoring using the following algorithms:

Recency scoring algorithm with input attribute cluster label, R values, and the output is R score:

1) Input the data set.
2) Group records by cluster label and find average R values in each cluster label.
3) Sort records by average R values from highest to lowest.
4) Generate new attribute R score from lowest to highest.

Frequency scoring algorithm with input attribute cluster label, F values, and the output is F score:

1) Input the data set.
2) Group records by cluster label and find average F values in each cluster label.
3) Sort records by average F values from lowest to highest.
4) Generate new attribute F score from lowest to highest.

Monetary scoring algorithm with input attribute cluster label, M values, and the output is M score:

1) Input the data set.
2) Group records by cluster label and find average M values in each cluster label.
3) Sort records by average M values from lowest to highest.
4) Generate new attribute M score from lowest to highest.

D. Evaluation

On this step, we evaluate each cluster performance produced on each layer by using the Davies-Bouldin Index (DBI). DBI is one of many methods used to evaluate the clustering algorithm. DBI is defined as the ratio of the sum of the within-cluster scatter to the inter-cluster separation and can be expressed as follows (5) [23], where the factor R_i can be written as (6) and S_i and S_j denote the within-cluster scatter for i_{th} and j_{th} clusters, respectively, and, e.g., S_i can be expressed as (7) where n_i is a number of x in the cluster C_i, and v_i is the center of this cluster. Moreover, the d_{ij} is the distance between the cluster centers, i.e., $d_{ij} = \| vi - vj \|$.

$$\mathrm{DB} = \frac{1}{\mathrm{K}} \sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{R_i} \quad (5)$$

$$\mathrm{R_i} = \max_{\mathrm{j} \neq \mathrm{i}} \frac{\mathrm{S_i} + \mathrm{S_j}}{\mathrm{d_{ij}}} \quad (6)$$

$$S_i = \frac{1}{n_i} \sum_{x \in C_i} ||x - v_i|| \quad (7)$$

The minimum of the DBI indicates the appropriate partitioning of a data set or in other words, the closer the value of DBI to zero indicates that the cluster produced are tight and well separated from each cluster center [24].

IV. EXPERIMENTS

For this experiment, we use RapidMiner Studio and as mentioned in the previous section, in this research we use a real dataset from a telecommunication company. Fig. 2 shows the step by step data processing starting from extracting the data from the database, data selection, attribute selection, RFM values generation, normalization, clustering using K-Means, evaluation using DBI and finally RFM score which is showing the results of segmentation.

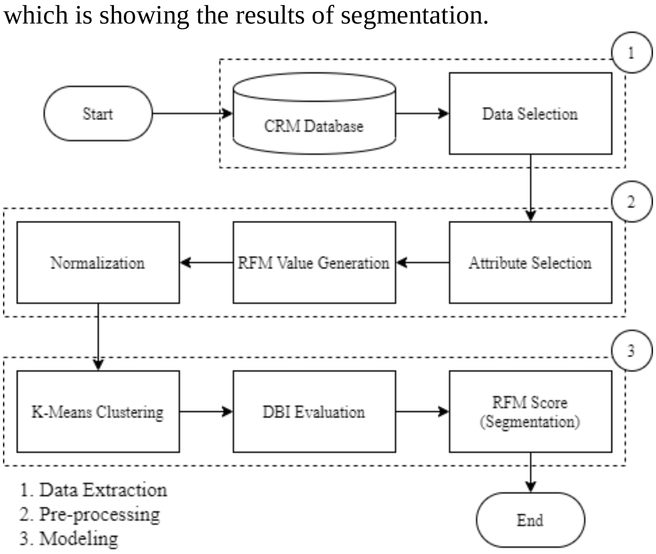

Fig. 2.Proposed Model

The data extracted from the CRM system for this study consist of 30838 records of an Active subscription status with 4 attributes. As mentioned in the previous section, we only choose the attributes that will be used to generate RFM values, so the attributes we chose are TRANS_ID, CUST_ID, MONTHLY_CHARGE, and START_DT as shown in Fig.3.

TRANS_ID	CUST_ID	MONTHLY_CHARGE	START_DT
0109201636	01-09-2016-4B7-3391	218531250	Sep 1, 2016 12:00:00 AM ICT
0109201637	01-09-2016-4B7-3391	218531250	Sep 1, 2016 12:00:00 AM ICT
0109201641	01-09-2016-4B7-3406	257239800	Sep 1, 2016 12:00:00 AM ICT
1805201746	18-05-2017-4F0-17	270000000	May 18, 2017 12:00:00 AM ICT
1805201747	18-05-2017-4F0-17	270000000	May 18, 2017 12:00:00 AM ICT
1304201748	13-04-2017-62L-290	275000000	Apr 13, 2017 12:00:00 AM ICT
2201201951	22-01-2019-1VXXET	333075000	Jan 22, 2019 12:00:00 AM ICT
2201201954	22-01-2019-1VXXET	355575000	Jan 22, 2019 12:00:00 AM ICT
0505201755	05-05-2017-2TF-3	365500000	May 5, 2017 12:00:00 AM ICT
0505201756	05-05-2017-4B7-276	365500000	May 5, 2017 12:00:00 AM ICT

Fig. 3.Data samples snapshot

After getting the Recency, Frequency and Monetary values in the manner mentioned in the previous section (pre-processing) then we left with 30360 records because the data has been grouped by CUST_ID. Now the data is ready to be normalized before processing it using the K-Means algorithm.

CUST_ID	R_VALUE	F_VALUE	M_VALUE ↓
01-01-2017-4B7-3401	37	1	35004960000
18-12-2017-4F0-17	26	2	12980500000
08-12-2017-4F0-14	26	2	12944750000
05-05-2017-2TF-3	33	1	12061500000
05-05-2017-4B7-276	33	1	12061500000
01-09-2016-4B7-3406	41	1	10546831800
13-04-2017-62L-290	34	1	9350000000
01-09-2016-4B7-3391	41	2	8959781250
22-01-2019-1VXXET	13	2	8952450000
18-05-2017-4F0-17	33	2	8910000000

Fig. 4.Generated Recency, Frequency and Monetary values from Fig. 3

After the data has been pre-processed, on this step we apply K-Means. Then we analyze each layer cluster performance using DBI as mentioned in the previous section (evaluation). From here we start by entering the values K = 2 to K = 10 and then comparing the DBI values to find the most optimal K value for each data layer.

Table- III: Davies-Bouldin Index result

K Value	R	F	M
2	-0.528	-0.322	-0.412
3	-0.479	-0.393	-0.319
4	-0.498	-0.380	-0.394
5	-0.478	-0.360	-0.412
6	-0.518	-0.415	-0.428
7	-0.544	-0.322	-0.444
8	-0.543	-0.284	-0.402
9	-0.522	-0.318	-0.409
10	-0.535	-0.320	-0.441

The result showed that on the R layer K = 5 is the most optimal K number with DBI score -0.478, on the F layer K = 8 is the most optimal K number with DBI score -0.318 and on the M layer K = 3 is the most optimal number with DBI score -0.319.

After finding the most optimal K value for each layer, now we set the RFM score based on the method we proposed in the previous section (clustering). For the R layer, the average value of each cluster has been sorted starting from the highest value to the lowest value based on the average value in each cluster and can be seen in Fig. 5. For the F layer and M layer, the average value of each cluster has been sorted starting from the lowest value to the highest value based on the average value in each cluster Fig. 6 and Fig. 7.

R_CLUSTER	average(RECENCY)	R_SCORE ↓
cluster_D	-1.662	5
cluster_E	-0.847	4
cluster_A	-0.164	3
cluster_B	0.479	2
cluster_C	1.136	1

Fig. 5.R layer cluster average value and R Score

F_CLUSTER	average(FREQUENCY)	F_SCORE ↓
cluster_F	66.262	8
cluster_E	51.923	7
cluster_B	33.999	6
cluster_G	21.452	5
cluster_D	11.453	4
cluster_H	7.113	3
cluster_C	3.528	2
cluster_A	-0.056	1

Fig. 6.F layer cluster average value and F Score

M_CLUSTER	average(MONETARY)	M_SCORE ↓
cluster_B	90.931	3
cluster_C	13.288	2
cluster_A	-0.035	1

Fig. 7.M layer cluster average value and M Score

At the final step, we determine values ranging from 1 to the number of all cluster labels in each cluster layer. In this research, the score on the R layer starts from 1 to 5, on the F layer the score starts from 1 to 8 and on the M layer the score starts from 1 to 3. After that, we join each RFM score to each record based on the respective CUST_ID as can be seen in Fig. 8.

CUST_ID	R_SCORE	F_SCORE	M_SCORE	R_VALUE	F_VALUE	M_VALUE ↓
01-01-2017-4B7-3401	4	1	3	37	1	35004960000
18-12-2017-4F0-17	4	2	2	26	2	12980500000
08-12-2017-4F0-14	4	2	2	26	2	12944750000
05-05-2017-2TF-3	4	1	2	33	1	12061500000
05-05-2017-4B7-276	4	1	2	33	1	12061500000
01-09-2016-4B7-3406	3	1	2	41	1	10546831800
13-04-2017-62L-290	4	1	2	34	1	9350000000
01-09-2016-4B7-3391	3	2	2	41	2	8959781250
22-01-2019-1VXXET	5	2	2	13	2	8952450000
18-05-2017-4F0-17	4	2	2	33	2	8910000000

Fig. 8.Generated RFM score for each customer

V. RESULTS AND DISCUSSION

With the proposed method in this research, it is proven that this method is superior in terms of simplicity and ease of use when compared to related research mentioned in the previous section, rather than adjusting the formula to generate RFM score, this scoring method can be done more dynamically when more data is available by changing the K values.

We have tested this model by trying several tests with different amounts of data, for data set 1 we are using 21587 records which are 70% of all the data, for data set 2 we use 27754 records which are 90% of all the data and finally data set 3 we use all 30838 records. the most optimal K value in each test always changes according to the amount of data. For further research, we want to develop this model to be able to use the most optimum K value automatically.

Table- IV: Davies-Bouldin Index result

Data Sets	Optimal K for R (score range)	Optimal K for F (score range)	Optimal K for M (score range)
Data Set 1	3	10	3
Data Set 2	5	9	3
Data Set 3	5	8	3

By doing this three-layer data clustering and producing an RFM score ranges that match the characteristics of the data by using optimal K value on each data layer, the company can divide the customers into different segments more comprehensively. This model shows the segment of each customer from several views. From each of these views, we can also find out which customers have the highest scores and lowest score.

For customers that have low RFM score, the company could plan appropriate marketing strategies to increase customer RFM values. For example, a customer F score is 1, and the frequency of that customer subscribing to service provided by the company is also 1, the company could plan marketing strategies by offering discounts or appropriate service bundling to increase F value of that customer. By increasing the F value of a customer, this also increases the M value of the customer and so on.

VI. CONCLUSION

The customer data in the CRM system can be massive and have hundreds to thousands of attributes. The method proposed in this study helps the enterprise in the telecommunication industry to identify customers segment based on each customer data. We have developed a novel method based on K-Means and DBI to generate RFM scores. Using this proposed method we can easily segment customers when the amount of data increases. Adjustment the number of the RFM score range was done by using the optimal number of K. Using this method we can determine the range of RFM scores objectively based on the amount of data available. Also, the proposed method could be implemented in other industries since it only depends on the values of RFM from the correspond data for each customer.

Now we have already completed this three-layer data clustering model based on RFM model, the marketing department in the company can create a marketing strategy that is right on target by their respective customer segments. This, of course, can improve the relationship between customers and companies because that is one of the objectives of implementing a CRM system.

Based on this model, we aim to develop a method where we can predict the customers RFM values in the future using historical transaction data.

APPENDIX

Table- V: Research Result

CUST_ID	R_SCORE	F_SCORE	M_SCORE	R_VALUE	F_VALUE	M_VALUE
01-09-2016-4B7-3391	3	2	2	41	2	8.95978125E9
01-09-2016-4B7-3406	3	1	2	41	1	1.05468318E10
01-01-2017-4B7-3401	4	1	3	37	1	3.500496E10
05-05-2017-2TF-3	4	1	2	33	1	1.20615E10
05-05-2017-4B7-276	4	1	2	33	1	1.20615E10
08-12-2017-4F0-14	4	2	2	26	2	1.294475E10
13-04-2017-62L-290	4	1	2	34	1	9.35E9
18-05-2017-4F0-17	4	2	2	33	2	8.91E9
18-12-2017-4F0-17	4	2	2	26	2	1.29805E10
22-01-2019-1VXXET	5	2	2	13	2	8.95245E9

ACKNOWLEDGMENT

The authors would like to express a sincere gratitude to the anonymous reviewers for their valuable comments and suggestions to improve the quality of this manuscript. In addition, the authors would also like to thank Bina Nusantara University for supporting this research project.

REFERENCES

1. M. H. R. S. M. Kamrul Islam Shaon, "A Theoretical Review of CRM Effects on Customer Satisfaction and Loyalty," Central European Business Review, vol. 4, no. 1, pp. 23-36, 2015.
2. S. K. Munyaradzi W. Nyadzayo, "The antecedents of customer loyalty: A moderated mediation model of customer relationship management quality and brand image," Journal of Retailing and Consumer Services, vol. 30, pp. 262-270, 2016.
3. R. R. C. O. Khamis Mwero Manero, "Customer Behaviour Segmentation among Mobile Service Providers in Kenya using K-Means Algorithm," International Journal of Computer Science Issues, vol. 15, no. 5, pp. 67-76, 2018.
4. J. V. Sivasankar Elango, "Customer Segmentation by Various Clustering Approaches and Building an Effective Hybrid Learning System on Churn Prediction Dataset," in Computational Intelligence in Data Mining, Singapore, 2017.
5. M. W. D. W. Dao Lam, "Clustering Data of Mixed Categorical and Numerical Type With Unsupervised Feature Learning," IEEE Access, vol. 3, pp. 1605-1613, 2015.
6. A. R. Azarnoush Ansari, "Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms," International Journal of Business and Management, vol. 11, no. 7, pp. 59-66, 2016.
7. R. D. A. M. K. P. J. M. N. J. du Toit, "Customer Segmentation Using Unsupervised Learning on Daily Energy Load Profiles," Journal of Advances in Information Technology, vol. 7, no. 2, pp. 69-75, 2016.
8. E. D. M. E. M. Mediana Aryuni, "Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering," in 2018 International Conference on Information Management and Technology (ICIMTech), Jakarta, 2018.
9. C. C. F. M. D. T. Ilias K. Savvas, "Understanding customers behaviour of telecommunication companies increasing the efficiency of clustering techniques," in 2017 25th Telecommunication Forum (TELFOR), Belgrade, 2017.
10. X. C. M. Y. C. S. L. X. H. Y. T. Z. Chen Cheng, "A novel cluster algorithm for telecom customer segmentation," in 2016 16th International Symposium on Communications and Information Technologies (ISCIT), Qingdao, 2016.
11. A. M, "Customer Segmentation Based on CLV Model and Neural Network," International Journal of Computer Science Issues (IJCSI), vol. 12, no. 2, pp. 31-37, 2016.
12. S. S. a. C. U. John Orange, "Using RFM segmentation to support alumni reunion fundraising," Journal of Education Advancement & Marketing, vol. 3, no. 4, pp. 325-339, 2019.
13. A. U. H. T. Peiman Alipour Sarvari, "Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis," Kybernetes, vol. 45, no. 7, pp. 1129-1157, 2016.
14. R. M. K. T. R. S. K. I. Hasitha Indika Arumawadu, "Mining Profitability of Telecommunication Customers Using K-Means Clustering," Journal of Data Analysis and Information Processing, vol. 3, no. 2, pp. 63-71, 2015.
15. T. M. Ahad Zare Ravasan, "A Fuzzy ANP Based Weighted RFM Model for Customer Segmentation in Auto Insurance Sector," International Journal of Information Systems in the Service Sector, vol. 7, no. 2, pp. 71-86, 2015.
16. Y.-R. H. H.-C. W. Chu Chai Henry Chan, "Marketing segmentation using the particle swarm optimization algorithm: a case study," Journal of Ambient Intelligence and Humanized Computing, vol. 7, no. 6, p. 855–863, 2016.
17. D. E. S. S. Nesma mahmoud Taher, "Investigation in Customer Value Segmentation Quality under Different Preprocessing Types of RFM Attributes," International Journal of Recent Contributions from Engineering, Science & IT (iJES), vol. 4, no. 4, p. 5–10, 2016.
18. N. S. G. A. M. Fariba Safari, "Customer lifetime value determination based on RFM model," Marketing Intelligence & Planning, vol. 34, no. 4, pp. 446-461, 2016.
19. M. A. Hadi Roshan, "The New Approach in Market Segmentation by Using RFM Model," Journal of Applied Research on Industrial Engineering, vol. 4, no. 4, pp. 259-267, 2017.
20. X. Z. H. E. Z. O. Meina Song, "Statistics-based CRM approach via time series segmenting RFM on large scale data," Knowledge-Based Systems, vol. 132, pp. 21-29, 2017.
21. V. C. M. V. D. V. S. S. P. K. D. Punya P Shetty, "Customers Churn Prediction with RFM Model and Building a Recommendation System using Semi-Supervised Learning in Retail Sector," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 1, pp. 3353-3358, 2019.
22. D. D. S. V. Preeti Aroraa, "Analysis of K-Means and K-Medoids Algorithm For Big Data," Procedia Computer Science, vol. 78, pp. 507-512, 2016.
23. A. Starczewski, "A new validity index for crisp clusters," Pattern Analysis and Applications, vol. 20, pp. 687-700, 2017.

24. M. R. R. S. A. Sadia Nawrin, "Exploring K-Means with Internal Validity Indexesfor Data Clustering in Traffic Management System," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 8, no. 3, pp. 264-272, 2017.G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

AUTHORS PROFILE

Afgan Fazri Handoko, has received his bachelor’s degree from Sekolah Tinggi Cipta Karya Informatika in Jakarta, Indonesia majoring in Computer Science in November 2016. He started working since July 2013 when also attending class for his bachelor’s degree and has worked for multiple Information Technology companies with his roles starting from software developer, system analyst, data analyst, business analyst and currently as a software consultant specialized in Customer Relationship Management. In September 2018, he entered. Graduate School of Bina Nusantara University (BINUS) in Computer Science Dept, majoring in Information Engineering. His research interest is in Machine Learning, Business Intelligence, Artificial Intelligence, and Big Data Analytics.

Antoni Wibowo, has received his first degree of Applied Mathematics in 1995 and a master degree of Computer Science in 2000. In 2003, He was awarded a Japanese Government Scholarship (Monbukagakusho) to attend Master and PhD programs at Systems and Information Engineering in University of Tsukuba-Japan. He completed the second master degree in 2006 and PhD degree in 2009, respectively. His PhD research focused on machine learning, operations research, multivariate statistical analysis and mathematical programming, especially in developing nonlinear robust regressions using statistical learning theory. He has worked from 1997 to 2010 as a researcher in the Agency for the Assessment and Application of Technology – Indonesia. From April 2010 – September 2014, he worked as a senior lecturer in the Department of Computer Science - Faculty of Computing, and a researcher in the Operation Business Intelligence (OBI) Research Group, Universiti Teknologi Malaysia (UTM) – Malaysia. From October 2014 – October 2016, he was an Associate Professor at Department of Decision Sciences, School of Quantitative Sciences in Universiti Utara Malaysia (UUM). Dr. Eng. Wibowo is currently working at Binus Graduate Program (Master in Computer Science) in Bina Nusantara University-Indonesia as a Specialist Lecturer and continues his research activities in machine learning, optimization, operations research, multivariate data analysis, data mining, computational intelligence and artificial intelligence.