JEST
TGO Journal of Education, Science and Technology
P-ISSN : 2987 - 4580
E-ISSN : 2987 - 4599
https://ejournal.trescode.org/index.php/jest

Savings Customer Segmentation Using RFMB Model and K-Means Clustering

Uus Firdaus[1], Azharudin[2], Alim Purnomo[3], Erlangga Gunawan[4]

Computer Science Study Program, Faculty of Computer Science, Universitas Djuanda[1,2,3,4]

Corresponding Author: Uus Firdaus (uus.firdaus@unida.ac.id)

Article Info

Received:
October 20, 2024
Revised:
November 10, 2024
Online available:
December 07, 2024

Keywords:
Segmentation, Customer, Banking, RFMB

Abstract

In increasing company growth, marketing strategies must be appropriate. Segmentation is carried out to create strategies by identifying potential customers. Recency Frequency Monetary (RFM) models analyze customer behavior. It is developed based on current transactions and balances and is called the RFMB model. Clustering can be interpreted as the process of grouping or classifying objects based on information obtained from data that can describe the relationships between objects. K-Means clustering algorithm that has the ability to group quite large amounts of data, and partition the dataset into several clusters. Base 65 thousand data from 1 branch, 147 thousand transactions in semester 1 2017, including cash withdrawals and deposits, book transfers, and ATM transactions. Results for 5 customer groups: SUPER, highest score for several attributes, 1,493 customers. PLATINUM, which has the highest balance, 44 customers. PREMIUM, 623 customers who have the largest transaction value. GOLD, 2,918 customers frequently transact, and CLASSIC's new transactions totaled 4062 customers.

Cite this as: Firdaus, U., Azharudin, Purnomo, A., & Gunawan, E. (2024). Savings Customer Segmentation Using RFMB Model and K-Means Clustering. TGO Journal of Education, Science and Technology, 2(2), 219–226.

INTRODUCTION

In efforts to increase the growth in the number of customers and products offered, one of them is savings products. Companies in the banking sector certainly need to implement good product marketing strategies so that they are targeted to their customers according to the characteristics of the customers and the products offered. Savings are designed to always be able to be withdrawn at any time by the owner through any media so that it is very possible to be used as material for segmenting these customers (Harun & Nathan, 2013).

In terms of supporting the company in its strategy to grow the number of customers,, as well as a marketing strategy so that it is right on target, this research is very important to carry out to get maximum results and make a contribution to the company and to academics. Product marketing, which is still carried out sporadically at various branch offices it has, will take time and money, especially the accuracy of customer data which is used as a reference in marketing its products. To meet the needs of business units, it is necessary to segment these savings customers.

Another thing, it will be very important to have a list of savings customers who in the near future will experience a change to passive because they rarely make transactions and require handling so that they remain connected, can make transactions and contribute to the company so that the company is helped to retain customers (Cheng & Chen, 2009).

Recency Frequency Monetary (RFM) model is an analytical method that focuses on habits or behavior that extracts customer profiles using several criteria and classifies customers using Recency (R), Frequency (F), and Monetary (M) attributes. R shows the recency of the last transaction, F is the number of transactions in a period, and M indicates the total amount of expenses in a period (Hu & Yeh, 2014).

In other research, there are other important attributes that can be considered for use in customer segmentation, namely the Balance (B) attribute. The balance attribute for customers is very important, because it will influence which customers are in the group and what kind of strategy will be carried out by the marketing department (Firdaus & Utama, 2020). Some articles improves the RFM concept by adding additional features or applying data mining techniques. R shows the time of last purchase, while F, M, and B show customer loyalty (Tavakoli, et al., 2018).

The method that can be categorized as state of the art in these segment customers using the RFM model, customers are given a rating for each attribute used from the lowest value to the highest. In line with the analysis process in the existing model, an idea to including the balance attribute in the savings account to determine which group the customer profile into based on transaction activity that has been carried out in a certain period of time.

Bank customers who are analyzed and segmented are customers for savings products, 1 (one) bank branch with a total of around 65 thousand data, customer transaction data that occurred in the first semester of 2017, namely in the period of January 1 2017 to 30 June 2017 with a total of around 147 thousand data. The scope of transaction data that occurs includes: cash withdrawals, cash deposits, book transfers, and transactions via ATM.

METHOD

The RFM model has been widely used in various applications involving a large number of customers such as online purchasing, retail, etc (Christy, Umamakeswari, Priyatharsini, & Neyaa, 2018). RFM model was proposed by Hughes in 1994 which differentiates important customers from big data by 3 (three) attributes: recency, frequency and monetary (Cheng & Chen, 2009).

R for recency of the last purchase refers to the interval between the time at which the last consumption behavior occurred and the current time. The shorter the interval, the greater the R value. F is related to the number of times a customer makes transactions. Frequency is defined as the number of purchases a customer makes in a certain period. The higher the frequency value, the more loyal the company's customers are. M relates to how much money customers transact. Monetary is defined as the amount of money spent by customers during a certain period. The higher the amount of money spent the more revenue they provide to the company.

Customer analysis using the RFM model can also be combined with other grouping methods in clustering algorithms in data mining (Hardiani, Sulistyo, & Hartanto, 2015). Studies conducted provide information that the greater the R, F values, the greater the likelihood that customers will generate new relationships with the company, and the greater M value, the greater the likelihood that customers will transact more of the company's services or products (Cheng & Chen, 2009).

Clustering is a grouping process that is categorized as unsupervised classification. Clustering can be interpreted as the process of grouping or classifying objects based on information obtained from data that can describe the relationships between objects. In data mining, clustering can be used to find distribution patterns in a dataset which functions for the data analysis process (Nainggolan & Lumbantoruan, 2018).

K-Means is a simple clustering algorithm that has the ability to group quite large amounts of data, and partition the dataset into k clusters. The K-Means algorithm has a high level of accuracy, is effective and requires relatively fast execution time because it is linear. The process of modifying the number k is carried out to obtain clusters whose members have a high level of similarity.

The steps of the K-Means algorithm can be described as follows:

1) Identify the data to be clustered and also determine the number of data clusters, Xij (i=1,2,....n; j= 1,2,....m), n data to be clustered and m the number data variables.
2) At the beginning of the iteration the center of each cluster is determined randomly (arbitrarily), Ckj (k = 1,.....k;j = 1,.....m).
3) Determine the distance of each data to the cluster center using the Euclidean formula.

$$d_{ij} = \sqrt{\sum_{j=1}^{m} (X_{ij} - C_{ij})2} \qquad (1)$$

$$C_{kj} = {}^{1}/_{n_j} \left(\sum\nolimits_{,} Z_p\right) \qquad (2)$$

To calculate the distance between vector data and each cluster centroid, use equation (1). Next, recalculate the cluster centroid using equation (2), and repeat this until it stops meeting the criteria.

Determining the number of k/clusters is the initial stage in the K-Means algorithm to obtain optimal grouping. To determine the best number of clusters, you can use the elbow method. The elbow method will provide the best cluster value taken from the Sum of Square Error (SSE) value.

The SSE value that experiences a significant decrease and that forms an elbow is considered optimal for determining the optimal number of clusters (Nainggolan & Lumbantoruan, 2018). Calculation of the SSE value can be done using the equation as shown in equation (3).

$$SSE = \sum_{i=1}^{k} d(p_i - m_i)^2 \qquad (3)$$

From equation (3), d is the minimum distance between the data and the cluster center point. Pi is the attribute of the ith data, and mi is the feature or attribute of the ith cluster center point.

RESULTS AND DISCUSSION

The model used to group savings customers based on customer attributes; time of last purchase (R), total number of transactions made (F), total nominal amount of money spent (M), as well as the customer's current balance attribute (Balance) are considered as one of the important attributes (Firdaus & Utama, 2020).

RFMB segmentation results are normalized for the grouping stage using the K-Means clustering method. The elbow principle is used to determine the optimal number of clusters and is evaluated with the Sum of Squared Errors (SSE).

Table 1. Identify segmentation attributes

Initial Attributes	Final Attributes
Last date of transaction (type: date)	Recency (type: number)
Number of transactions	Frequency
The accumulated amount of transactions	Monetary
The total balance in the account	Balance

Tabel 2. Determination of parameter values for each segmentation attribute

Nilai Atribut	Nilai			
	1	2	3	4
Recency	> 50	31-50	16-30	0-15
Frequency	<= 6x	7-24	25-72	> 72
Monetary	<= 600.000	600.001 - 6.000.000	6.000.001 - 24.000.000	> 24.000.000
Balance	<= 1.000.000	1.000.001 - 10.000.000	10.000.001 - 100.000.000	> 100.000.000

Next, determine the score value for each customer. To get a score from each customer, the R, F, M and B values are added up, so that the formula for the Total Score is as shown in equation (4).

$$Total\ Score = R + F + M + B \tag{4}$$

The segment of each customer can be seen from the segment code and from the Total_score of that customer. Total_Score can describe where the customer is, the same Total_score that occupies that segment. The segmentation data is normalized first so that the clustering process can then be continued.

The process begins by determining the number of clusters and determining the number of iterations that can be carried out randomly. Determining the optimal number of clusters can use the elbow method and SSE calculation in equation (5), while calculating the distance of each data to the cluster center with equation (2). Repeat the cluster centroid calculation with equation (3) until all data is processed.

Each variable is dependent on filling in the data and processes carried out by other variables. The recency component variable depends on the input that comes from inputting the last date of the transaction, the frequency variable comes from inputting the total number of transactions, the monetary variable comes from inputting the total nominal transaction, while the balance variable comes from the system component, namely the customer balance. The output variable depends on the combination of several system component variables such as recency, frequency, monetary, balance, where the combined results of these variables produce segmentation of savings customers based on their supporting attributes.

In the integration data, it will be shown that there are several fields that are used as segmentation attributes in the previous discussion regarding the identification of segmentation attributes. Next, these attributes are given assessment parameters to carry out the calculation and accumulation process for each customer. These parameter values are shown in Table 2.

Based on the integration data that has been calculated and accumulated for each customer, segmentation will then be carried out using predetermined value parameters so that each customer will get a value for each attribute (R, F, M, B) and get a segmentation code that shows the profile. customers based on these values. The calculated and accumulated data for each customer is as shown in Table 3, and the segmentation data for each customer is shown in Table 4.

Tabel 3. Calculation results for the Recency, Frequency, Monetary, Balance attributes

Account	Recency	Frequency	Monetary	Balance
10105****	9	72	112.083.706	9.157.730
10108****	10	90	42.532.143	1.571.210
10109****	6	76	47.246.700	17.645.800
10110****	89	1	5.000.000	2.066.250
10110****	59	3	1.675.000	141.684
10110****	8	74	217.936.797	12.987.100
10110****	9	31	267.226.500	23.006.600
10110****	179	1	1.400.000	44.085
10110****	2	58	91.601.100	81.927

Table 4. Segmentation Data Based on RFMB Model

Account	LastTransDate	Recency	Frequency	Monetary	Balance	R	F	M	B
10105****	22-Jun-2017	9	72	112.083.706	9.157.730	4	3	4	2
10108****	21-Jun-2017	10	90	42.532.143	1.571.210	4	4	4	2
10109****	25-Jun-2017	6	76	47.246.700	17.645.800	4	4	4	3
10110****	3-Apr-2017	89	1	5.000.000	2.066.250	1	1	2	2
10110****	3-May-2017	59	3	1.675.000	141.684	1	1	2	1
10110****	23-Jun-2017	8	74	217.936.797	12.987.100	4	4	4	3

To make it easier to identify customers in each segment, total_score is used as the label for that customer segment. Table 5 is snapshot of segment data from total_score.

Table 5. Snapshot of customer segment data based on total score

Account	R	F	M	B	Total_Score
10105****	4	3	4	2	13
10108****	4	4	4	2	14
10109****	4	4	4	3	15
10110****	1	1	2	2	6
10110****	1	1	2	1	5
10110****	4	4	4	3	15
10110****	4	3	4	3	14
10110****	1	1	2	1	5
10110****	4	3	4	1	12
10111****	4	3	4	2	13

Table 6. Customer Segmentation Results

Group	Number of member
SUPER	1.493
PLATINUM	44
PREMIUM	623
GOLD	2.918
CLASSIC	4.062

Segmentation results are defined and represented into customer groups which are mappings from savings customer data, resulting in 5 customer groups, namely:

1. SUPER, having the highest combination of values for each attribute (Recency=”New”, Frequency=”Frequently”, Monetary=”Many”, Balance=”Many”), with a total of 1493 customers.
2. PLATINUM, having the highest B value (Balance = "Many") regardless of the other attribute values. This group consists of 44 customers.
3. PREMIUM, having the highest M value (Monetary = "A lot") regardless of the other attribute values. A total of 623 customers are in this group.
4. GOLD, having the highest F value (Frequency = "Frequently") regardless of the other attribute values. This group consists of 2918 customers.
5. CLASSIC, have the highest R value (Recency = "New") regardless of the other attribute values. This group totals 4062 customers.

Next process is the clustering process of the segmentation data. Grouping uses a clustering algorithm to obtain groups that have similarities between their members. The main concept in the clustering method that is emphasized is the search for the optimal number of clusters. For this reason, the elbow method is used by obtaining and paying attention to the Sum of Squared Error (SSE) value. SSE states the total sum of the squared values of the data distance to the cluster center. The smaller the SSE value, the better the results (Qi, Yu, Wang, & Liu, 2016). The SSE equation will get the number of clusters that will be used in the K-Means algorithm. In Figure 1 it can see the position that forms an elbow, the curve before it becomes sloping is located at the number 5, so this value is applied to the clustering process with the number of groups being 5.

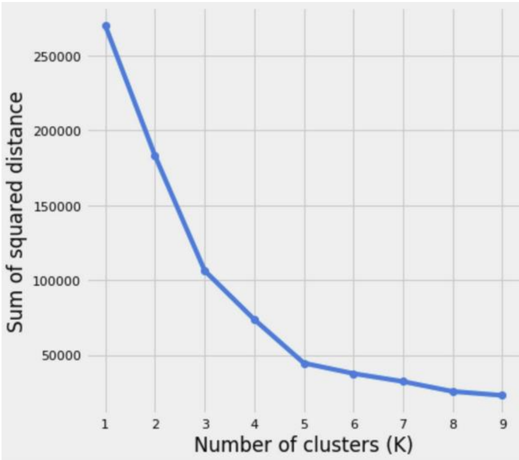

Figure 1. Elbow Method - Number of groups used

Table 7. Data snippet for degree of membership of the clustering algorithm

Data Ke-	Derajat keanggotaan					Jarak Minimum	Cluster				
	d_i1	d_i2	d_i3	d_i4	d_i5		1	2	3	4	5
1	1,4346	2,8099	1,1407	3,1432	0,3429	0,3429	0	0	0	0	1
2	1,2158	1,0326	1,0449	3,6582	2,3118	1,0326	0	1	0	0	0
3	1,7050	2,5059	1,2441	3,9859	1,0205	1,0205	0	0	0	0	1
4	1,6786	2,8459	2,0294	3,9954	1,0549	1,0549	0	0	0	0	1
5	1,0601	1,9130	3,1242	3,1605	1,2669	1,0601	1	0	0	0	0
6	1,8788	1,5151	1,5007	3,7977	3,1104	1,5007	0	0	1	0	0
7	1,4970	1,2007	1,1618	3,0055	3,8974	1,1618	0	0	1	0	0
8	1,4005	3,9714	3,9816	0,3692	3,6805	0,3692	0	0	0	1	0
9	1,9471	3,3497	1,2350	0,8653	1,8486	0,8653	0	0	0	1	0
10	3,1898	1,2014	3,2880	2,2135	1,2690	1,2014	0	1	0	0	0

The degree of membership and tendency to be in a cluster also has similarities between grouping using the K-Means clustering algorithms. Example data for calculating membership degrees in the K-Means algorithm is shown in Table 7, and Figure 2 shows the level of accuracy obtained from grouping using 5 (five) clusters of 77.58%.

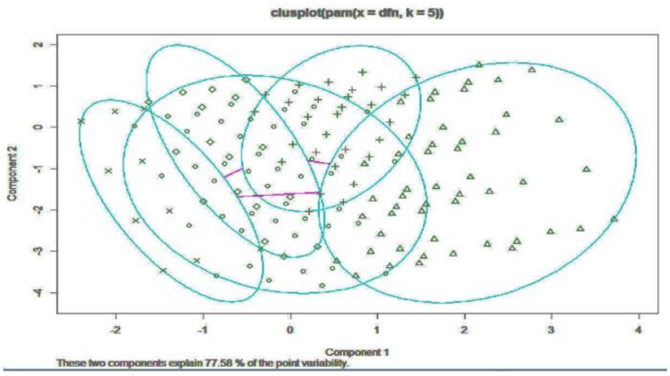

Figure 2. Plot to determine the level of grouping accuracy

CONCLUSION

Based on the analysis and trials in this research, conclusions can be drawn:

1. The model used can be adapted to research needs and in an effort to resolve problems found in the research company related to savings customer segmentation based on company customer transactions and data.
2. This model includes the balance attribute as an attribute owned by the company for segmentation and adapts to the research objectives. The RFMB model is expected to become an analytical method and reference model for research related to segmentation.
3. The customer grouping model in this research was carried out using the K-Means clustering algorithm. From the processed company data, it has succeeded in dividing savings customers into 5 groups and has been validated and confirmed by the business unit for the accuracy of the results of this customer grouping.
4. Segmentation result groups, SUPER with the highest combination of scores has 1493 customer, PLATINUM with the highest balance value has 44 customers, PREMIUM with the largest number of transaction has 623 customers, GOLD totaling 2918 customers who are classified as the group of customers who make transactions most frequently, and CLASSIC is a group who have carried out transactions has 4062 customers

REFERENCES

Cheng, C.-H., & Chen, Y.-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications, 4176–4184. https://doi.org/10.1016/j.eswa.2008.04.003.

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2018). RFM Ranking – An Effective Approach to Customer Segmentation. Journal of King Saud University - Computer and Information, https://doi.org/10.1016/j.jksuci.2018.09.004.

Firdaus, U., & Utama, D. N. (2020, 05). Balance as one of the attributes in the customer segmentation analysis method: Systematic Literature Review. Advances in Science, Technology and Engineering Systems Journal, 5(3), 334-339.

Hardiani, T., Sulistyo, S., & Hartanto, R. (2015). Segmentasi Nasabah Tabungan Menggunakan Model RFM (Recency, Frequency,Monetary) dan K-Means Pada Lembaga Keuangan Mikro. Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK) (hal. 463-468). ISBN: 979-26-0280-1.

Harun, C. A., & Nattan, R. R. (2013). Non Core Deposit: Studi terhadap Dana Pihak Ketiga Perbankan Indonesia. Jakarta: Bank Indonesia.

Hu, Y.-H., & Yeh, T.-W. (2014). Discovering valuable frequent patterns based on RFM analysis without. Knowledge-Based Systems, 76-88. https://doi.org/10.1016/j.knosys.2014.02.009.

Mawoli, M. A. (2012). Effective Market Segmentation and Viability of Islamic Banking in Nigeria. Australian Journal of Business and Management Research Vol.1, 1-9.

Nainggolan, R., & Lumbantoruan, G. (2018). Optimasi performa cluster K-means menggunakan Sum of Squared Errors (SSE). METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi Vol. 2 No.2, 103-108. ISSN: 2620-4339.

Newell, F. (1997). The new rules of marketing: How to use one-to-one relationship. New York: McGraw-Hills Companies.

Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., & Rahmani, R. (2018). Customer Segmentation and Strategy Development based on User Behavior Analysis, RFM model and Data Mining Techniques: A Case Study. IEEE. International Conference on e-Business Engineering (ICEBE), (hal. 119-126. DOI 10.1109/ICEBE.2018.00027)