JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 3, No. 4 September 2021 DOI: https://doi.org/10.34288/jri.v3i4.98

COMPARATIVE STUDY OF FUZZY C-MEANS AND K-MEANS ALGORITHM FOR GROUPING CUSTOMER POTENTIAL IN BRAND LIMBACK

Difa Lazuardi Aditya [1*], Devi Fitrianah [2]

Teknik Informatika
Universitas Mercu Buana, Jakarta, Indonesia
https://www.mercubuana.ac.id
difa.lazuard@gmail.com [1*], devi.fitrianah@mercubuana.ac.id [2]
(*) Corresponding Author

Abstrak

Pelanggan adalah suatu pemangku kepentingan bagi sebuah bisnis, untuk menjaga dan meningkatkan antusias pelanggan serta mengembangkannya terhadap kinerja perusahaan maka perlu dilakukannya segmentasi pelanggan yang bertujuan untuk mengetahui pelanggan potensial. Pada penelitian ini menggunakan data transaksi pembelian dari pelanggan Brand Limback pada periode tahun 2020. Penggunaan analisis RFM (Recency, Frecuency, Monetary) membantu dalam menentukan atribut yang di gunakan untuk segmentasi pelanggan. Untuk menentukan jumlah cluster yang optimal dari dataset RFM maka di terapkannya metode Elbow. Dataset yang di hasilkan dari RFM di kelompokkan menggunakan algoritma Fuzzy C-Means dan K-Means, dari kedua algoritma akan di bandingkan kualitas dalam pembentukkan clusternya menggunakan metode Silhoutte Coefficient dan Davies-Bouldin Index. Segmentasi pelanggan dari dataset RFM yang telah di cluster menghasilkan 7 cluster yang optimal yaitu Cluster 0 merupakan bronze customer. Cluster 1 merupakan silver customer. Cluster 2 merupakan gold customer. Cluster 3 merupakan platinum customer. Cluster 4 merupakan diamond customer. Cluster 5 merupakan super customer, dan cluster 6 merupakan superstar customer. Validasi cluster dari k-means menggunakan silhouette coefficient menghasilkan nilai 0.934 sedangkan davies bouldin index menghasilkan nilai 0.155 dan hasil validasi algoritma fuzzy c-means menggunakan silhouette coefficient menghasilkan nilai 0.921 sedangkan davies bouldin index menghasilkan nilai 0.145.

Kata kunci: Clustering, Segmentasi Pelanggan, RFM, Fuzzy C-Means, K-Means

Abstract

The customer is a stakeholder for a business, to maintain and increase customer enthusiasm and develop it for the company's performance, it is necessary to do customer segmentation which aims to find out potential customers. This study uses purchase transaction data from Brand Limback customers in the period 2020. The use of RFM (Recency, Frecuency, Monetary) analysis helps in determining the attributes used for customer segmentation. To determine the optimal number of clusters from the RFM dataset, the Elbow method is applied. The datasets generated from RFM are grouped using the Fuzzy C-Means and K-Means algorithms, the two algorithms will compare the quality in the formation of clusters using the Silhoutte Coefficient and Davies-Bouldin Index methods. Customer segmentation from the RFM dataset that has been clustered produces 7 optimal clusters, namely Cluster 0 is a bronze customer. Cluster 1 is a silver customer. Cluster 2 is a gold customer. Cluster 3 is a platinum customer. Cluster 4 is a diamond customer. Cluster 5 is a super customer, and cluster 6 is a superstar customer. The cluster validation of k-means using the silhouette coefficient produces a value of 0.934 while the Davies bouldin index produces a value of 0.155 and the validation results of the fuzzy c-means algorithm using the silhouette coefficient produces a value of 0.921 while the Davies bouldin index produces a value of 0.145.

Keywords: Clustering, Customer Segmentation, RFM, Fuzzy C-Means, K-Means

INTRODUCTION

Rapid technological advances have given rise to many marketplaces or platforms that function as markets or third parties between sellers and buyers to facilitate online buying and selling transactions. Every marketplace has an operating system in every business transaction that is always recorded and written down. The data is stored in a large capacity database. For companies, the data stored in the database can be used to create sales reports, inventory control, and so on.

According to a survey released by bps.go.id on e-commerce 2020 statistics, there is an increase every year for new businesses operating. According to records, between 2010-2016 as many as 38.58% of businesses have started operating. Furthermore, in 2017-2019 new businesses operating increased to 45.93% (A. Luthfi, N. Anggraini, 2020). With a very rapid increase in business people from year to year, it will certainly have an impact on quite tight competition. The main focus of the company's competition with competitors is the customer. Customers play an important role in the development of business strategies, and customers are also a source of company profits. Therefore, companies need a full understanding of customers. Many companies have difficulty identifying suitable customers, this is something that can cause companies to lose potential customers, which of course will cause great damage to the company (Nurmalasari et al., 2020). For this reason, companies are required to get to know their customers better in order to maintain loyalty to the company by segmenting customers.

Research conducted by Stephen et al (Sutresno, Iriani, & Sediyono, 2018), applies the K-Means clustering method and RFM model. The object of research used is customer data in one of the workshops during the period 2015 to 2017 which was obtained through observations and interviews with one of the workshop managers and obtained 15,913 customer data, 40,361 transaction data and 76,214 detailed transaction data. The results obtained from the application of K-Means clustering with the RFM model on customer transaction data in the workshop are placing 14,979 customer data groups in 5 clusters.

Another study conducted by Bena et al (Ashari, Otniel, & Rianto, 2019), discusses the comparison of the performance of K-Means with DBSCAN for clustering sales data. The cluster results obtained from the two algorithms are 3 clusters, where in K-Means cluster 1 is data that has a large number of sales per transaction and a high price of goods, for cluster 2 is data that has a small number of sales per transaction and a low price and for cluster 3 is data where the number of sales per transaction is small and the price is high. As for the results from DBSCAN, cluster 1 is data with high prices, cluster is data with medium prices and cluster 3 is data with low prices. The remaining 74 data are noise that does not fit into any cluster.

The purpose of this study is to compare the K-Means and Fuzzy C-Means algorithms in determining customer segmentation which will then be used to identify potential customers using RFM models (Recency, Frequency, Monetary) to make it easier to determine efficient attributes from the many existing attributes. on the previous data, then the results of the RFM data will be clustered using the fuzzy c-means algorithm and the k-means algorithm.

RESEARCH METHODS

This study uses a quantitative approach where the type of research is based on data that already exists and is ready to be processed. In this study, the author uses python tools and also conducts experiments by comparing the performance of the K-Means and Fuzzy C-Means algorithms to find out which algorithm has the best level of accuracy. In this study, the authors perform several stages. The stages of the research can be seen in Figure 1 below:

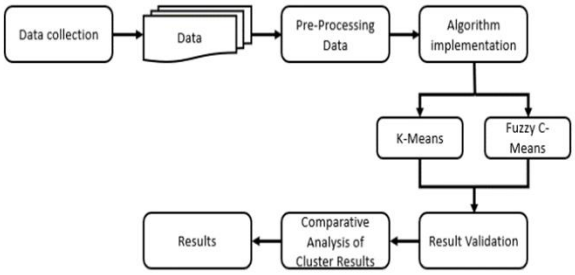

Figure 1. Research Stages

Data Mining

Data mining is an iterative and interactive process to find new patterns or models that are perfect, useful and understandable in a very large database. Data mining contains the search for the desired trend or pattern in a large database to help make decisions in the future (Syahdan & Sindar, 2018).

Data mining uses statistical techniques, artificial intelligence and machine learning to extract and identify information from various large databases in its processing (Ashari et al., 2019).

Cluster Analysis

Cluster analysis is a statistical technique that aims to group objects into a group so that objects in one group have a higher similarity than objects in other groups. In other words, the purpose of cluster analysis is to classify objects based on their similarity and collect data into several groups (Silvi, 2018).

Cluster analysis uses multivariate techniques to group objects with similar characteristics (Sadewo, Windarto, & Wanto, 2018).

Data collection

The data used in this study was obtained from the Brand Limback database on Tokopedia in the

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v3i4.98

2020 period, which amounted to 2,226 transactions.

Pre-Processing Data

Before entering the analysis stage on the dataset, the data preprocessing stage must be carried out first which aims to reduce errors at the analysis stage, because sometimes the dataset experiences data loss or errors when entering data (Kiat, Azhar, & Rahmayanti, 2020). Preprocessing is the process of changing data to be simpler and more effective as needed (Saifullah, Zarlis, Zakaria, & Sembiring, 2017). Data preprocessing carried out in this study includes:

1. Data Cleansing, is the process of deleting invalid data or the same data in one column.
2. Data Selection, is the selection of attributes from the main dataset that will be used as a new dataset to be processed at the grouping stage. In this study, the data selection process uses the RFM model which consists of Recency, the time span of the transaction. Frequency, total transactions. Monetary, the total cost of the total transaction. That way we will get a customer dataset with a simple dataset with attributes: kdcustomer, recency, frequency, monetary.
3. Min-Max Scaler, is a linear transformation process or assigns a scale value from 0 to 1 to the original data to be used (Wiranda & Sadikin, 2019).

Elbow Method

Elbow method is a method that aims to help find the optimal number of clusters in the dataset. To determine the optimal K value, the K value will be checked one by one and the SSE (Sum Square Error) value will be recorded. SSE (Sum of Square Error) is a formula used to measure the difference between the data obtained with the forecast model that has been done previously (Winarta & Kurniawan, 2021).

RFM Model

RFM is an analytical method to identify customer attitudes and represent customer attitudes based on 3 attributes, namely Recency, Frequency, Monetary.

According to Tsiptsis and Chorianopoulos (2009), RFM analysis consists of Recency, Frequency, Monetary which has the following meanings (Taqwim, Setiawan, & Bachtiar, 2019):

1. Recency, is a variable to measure customer value based on the time span (date, month, year) of the customer's last transaction to date. The smaller the time span, the greater the recency value.
2. Frequency, is a variable to measure customer value based on the number of transactions made by customers in one period. The greater the number of transactions carried out, the greater the f value.
3. Monetary, is a variable to measure customer value based on the amount of money issued by customers in one period. The greater the amount of money spent by the customer, the greater the M value.

The RFM model is a popular method and is often used to analyze customer behavior as seen from the combination of the three-digit value of each RFM attribute, where each attribute has a value based on quintiles with intervals of 1 to 5 which are evenly distributed. The value assigned to each attribute represents the good or bad score of the customer, a value of 5 is said to be the best and 1 is the worst which will then form a combination of numbers, such as 555, 554, 553, ..., 111 (Sutresno et al., 2018).

K-Means Algorithm

K-Means is a non-hierarchical data grouping method that divides data into two or more groups. This method divides data into several groups so that data with the same characteristics are included in the same group, and data with different characteristics are grouped into other groups. The objective function used for K-Means is determined based on the distance and value of the data objects in the group (Putra & Wadisman, 2018). The steps in performing clustering with k-means are as follows (Triyansyah & Fitrianah, 2018):

1. Specifies the number of clusters.
2. Allocate data randomly into existing clusters according to the closest distance.
3. Calculate the average of each cluster from the data in each cluster.
4. Re-allocate all data to the cluster according to the closest distance.
5. Repeat process number 3, until no changes or changes occur.

Fuzzy C-Means Algorithm

Fuzzy C-Means is one of the fuzzy clustering methods which was first developed by (Dewangan & Ambhaikar, 2013) and later modified by (Dulyakarn & Rangsanseri, 2001) as a method that is often used in pattern recognition. The FCM grouping of data is based on membership levels between 0 and 1, this causes grouping of data that not only has the same value in a cluster, but also groups of values that have two or more groups according to their membership level (Yunita, Herman, Takwim, & Widianto, 2019). The stages of the FCM algorithm are as follows (Klawonn, 2007 in

Sumanto and Wahono, 2011) (Rustiyan & Mustakim, 2018):

1. Identify the data to be clustered in the form of a matrix with a size of n x m (x_{ij}) is the i-th sample data (i = 1, 2, n) and the j-th attribute (j = 1, 2, m).
2. Define:
   a. Number of Clusters : c;
   b. Rank : w;
   c. Maximum Iterations : MaxIter;
   d. Error Rate : I;
   e. Initial Objective Function : P^o = 0;
   f. Early Iteration : t = 1;
3. Generate random number μik, i=1,2...n; k=1,2...c; as elements of the initial partition matrix U

$$Qi = \sum_{k}^{c} = 1\mu ik \quad (1)$$

With j=1,2...n
Count:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (2)$$

4. Calculate the value of the k-th cluster center (v_{kj}) where k=1,2,...c and j=1,2,...m) using the formula:

$$v_{kj} = (\mu_{ik})^w \Sigma_{j=1}^{m} (\mu_{ik})^m \quad (3)$$

5. Calculate the objective function at the t-th iteration (pt):

$$pt = \Sigma_{i=1}^{n} \Sigma_{k=1}^{c} \left(\left[\Sigma_{j=1}^{m} (x_{ij} - v_{kj})^2 \right] (\mu_{ik})^w \right) \quad (4)$$

6. Calculate the change in the partition matrix u:

$$\mu_{ik} = \frac{\left[\Sigma_{j=1}^{m} (x_{ij} - v_{kj})^2\right] \frac{-1}{w-1}}{\Sigma_{k=1}^{c} \left[\Sigma_{j=1}^{m} (x_{ij} - v_{kj})^2\right] \frac{-1}{w-1}} \quad (5)$$

7. Check stop condition:
   a. If (pt–pt–1)< Ɛ or (t<maxiter), then the iteration stops.
   b. If t = t +1, repeat step-4 to step-6.

Evaluation and Validity

Validation is the testing phase of the cluster that has been obtained to find out how well the performance of the cluster is. In this study, the authors propose 2 validations, namely, the Silhouette Coefficient and the Davies Bouldin Index.

Silhouette Coefficient method is one of the methods used to test the quality of clusters from the clustering process. Silhouette Index will evaluate the placement of each object in each cluster by comparing the average distance of objects in one cluster and the distance between objects in different clusters (Nahdliyah, Widiharih, & Prahutama, 2019). The following is the equation used in the global silhouette (Adiana, Soesanti, & Permanasari, 2018):

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

The explanation is as follows, $S(i)$ is the silhouette value. $b(i)$ is the average distance from object i to all objects in the same cluster. $a(i)$ is the average distance from object I to objects in different clusters. Based on the standardized silhouette, where a higher value is better than a lower value, a value close to zero is considered not good (Christina Deni Rumiarti, 2017).

The Davies Bouldin Index (DBI) method was first proposed in 1979 by David L. Davies and Donald W. Bouldin. DBI evaluation is seen from the number and proximity of data from clustering results, where whether or not the cluster results are seen from the quantity and proximity between the data from the cluster results. The DBI measurement method is to maximize the distance between clusters and minimize the distance between clusters (Jollyta, Efendi, Zarlis, & Mawengkang, 2019). The smaller the DBI value, the better the cluster. DBI value is formulated as follows (Tempola, Muhammad, & Mubarak, 2020):

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i=j} (R_{i,j}) \quad (7)$$

Where K is the number of clusters used. $R_{i,j}$ is the ratio of the comparison between the i-th cluster and the j-th cluster. In DBI validation, a cluster can be said to be good if it has the smallest possible cohesion and as much separation as possible. $R_{i,j}$ formulated as follows:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (8)$$

For SSW (Sum of Square within cluster) it is a cohesion metric in a cluster I, as follows:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d\,(x_j, c_i) \quad (9)$$

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v3i4.98

Where m_i is the number of data in the i-th cluster. c_i is the i-th cluster centroid. $d()$ the same as the euclidean distance.

RESULTS AND DISCUSSION

After the process of data cleansing and data selection using the RFM model, as can be seen in Table 2, is the dataset that will be used for clustering the K-Means and Fuzzy C-Means algorithms. From the results of preprocessing, a new dataset of 2,207 transactions was obtained from the initial data which previously amounted to 2,226 transactions in the 2020 period.

Pre-Processing Data

At the data preprocessing stage there are several stages that are carried out. The first step is data cleansing. The second step is data selection. The third step is the min-max scaler.

In the data cleansing stage, invalid data or the same data in one column will be deleted, the results of data cleansing can be seen in Table 1.

Table 1. Data Cleansing Results

OrderID	Invoice	Payment Date	Order Status
590032536	INV/20200919/XX/IX/631592429	19/09/2020 10:48	Transaksi selesai Dana diteruskan ke penjual.
590433824	INV/20200919/XX/IX/631993717	19/09/2020 22:24	Transaksi dibatalkan.
591481423	INV/20200919/XX/IX/633041316	21/09/2020 11:33	Transaksi selesai Dana diteruskan ke penjual.

The next step is data selection or dataset selection from the main data, the data selection process uses the RFM model, the results of which can be seen in Table 2.

Table 2. Data Selection Results

OrderID	Recency	Frequency	Monetary
138230982	358	1	159900
138253556	358	1	200000
138317446	358	1	159900
138389674	358	1	125000
138667638	358	1	69000
...	...	...	...

After the new dataset is obtained, the next step is the min-max scaler to change the data range to 0 to 1, the results of which can be seen in Table 3.

Table 3. Min-Max Scaler Results

OrderID	Recency	Frequency	Monetary
1.00000000e+00 1.20151501e-01	1.00000000e+00 1.20151501e-01	1.00000000e+00 1.20151501e-01	1.00000000e+00 1.20151501e-01
1.00000000e+00 7.08162221e-02	1.00000000e+00 7.08162221e-02	1.00000000e+00 7.08162221e-02	1.00000000e+00 7.08162221e-02
9.99958060e-01	9.99958060e-01	9.99958060e-01	9.99958060e-01
1.41276947e-01	1.41276947e-01	1.41276947e-01	1.41276947e-01
9.99803414e-01	9.99803414e-01	9.99803414e-01	9.99803414e-01
6.93417525e-02	6.93417525e-02	6.93417525e-02	6.93417525e-02
...	...	...	...

Elbow Method Analysis

Before entering the clustering stage, the K value will be determined using the elbow method to obtain the optimal K value. Figure 2 shows a graph of the results of the Elbow method, the optimal K value can be seen from the significant line changes on the graph. So it can be concluded that the optimal number of clusters from the RFM dataset that has been generated according to Elbow is 7 clusters.

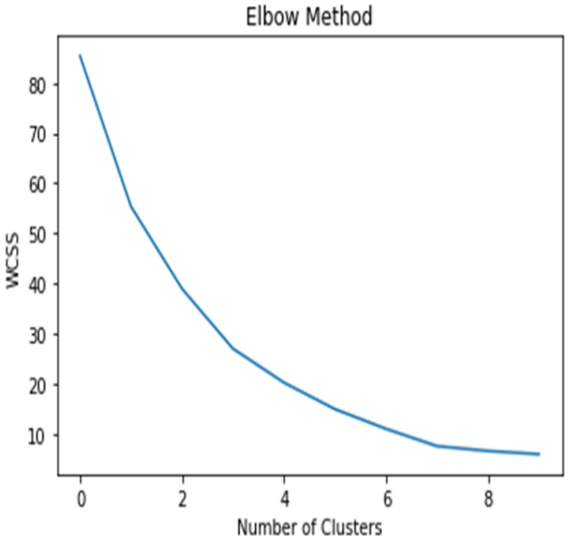

Figure 2. Elbow Method

K-Means Cluster Analysis

The results of the K value obtained from the elbow method will be used as the number of test clusters with a value of K = 7 and the results of cluster validation from k-means using the Silhouette Coefficient and Davies Bouldin index can be seen in Table 4.

Table 4. K-Means Validation Results

Algorithm	K	Silhouette Coefficient	Davies Bouldin Index
K-Means	7	0.93	0.15

It can be seen in table 4 that the k-means algorithm gets the highest validation results with the silhouette coefficient. The visualization results of the k-means cluster can be seen in Figure 3.

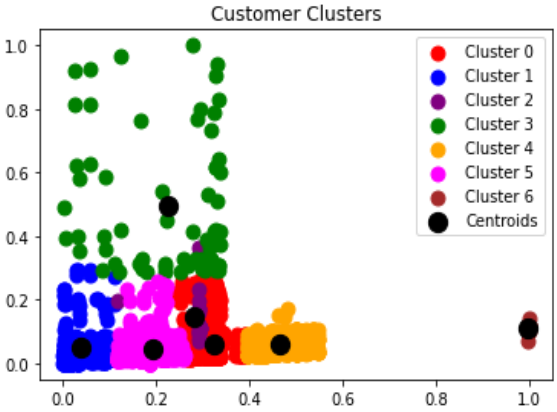

Figure 3. K-Means Cluster Visualization Results

Figure 3 shows that cluster 0 consists of 522 customers, which have an RFM value range of 311-515. Cluster 1 consists of 355 customers, which have an RFM value range of 111-115. Cluster 2 consists of 352 customers, which have an RFM value range of 511-515. Cluster 3 consists of 67 customers, which have an RFM value range of 215-515. Cluster 4 consists of 19 customers, which have an RFM value range of 354-555. Cluster 5 consists of 884 customers, which have an RFM value range of 111-512 and cluster 6 consists of 8 customers, which have an RFM value range of 112-115.

Fuzzy C-Means Cluster Analysis

In fuzzy c-means, the number of test clusters used is still the same as the K-Means, which is 7 clusters, with the aim of comparing the results of the cluster patterns formed. The results of the cluster validation of Fuzzy C-Means using the Silhouette Coefficient and Davies Bouldin index can be seen in Table 5.

Table 5. Fuzzy C-Means Validation Results

Algorithm	K	Silhouette Coefficient	Davies Bouldin Index
Fuzzy C-Means	7	0.92	0.14

It can be seen in table 5 that the Fuzzy C-Means algorithm gets the highest validation results with the Davies Bouldin Index. The visualization results of the K-Means cluster can be seen in Figure 4.

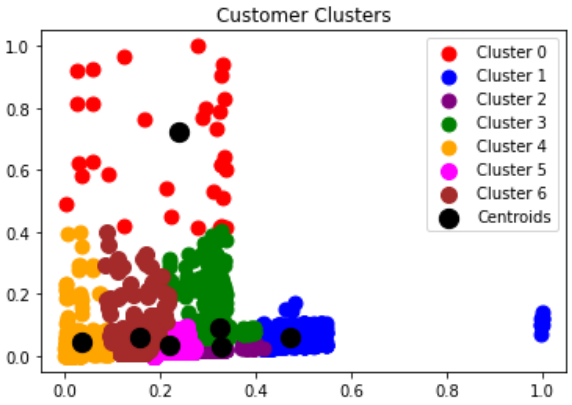

Figure 4. Fuzzy C-Means Cluster Visualization Results

Figure 4 shows that cluster 0 consists of 33 customers, which have an RFM value range of 215-515. Cluster 1 consists of 345 customers, which have an RFM value range of 111-115. Cluster 2 consists of 481 customers, which have an RFM value range of 111-411. Cluster 3 consists of 453 customers, which have an RFM value range of 114-415. Cluster 4 consists of 339 customers, which have an RFM value range of 511-515. Cluster 5 consists of 327 customers, which have an RFM value range of 311-415 and cluster 6 consists of 229 customers, which have an RFM value range of 411-515.

Validation Analysis

Based on the parameters used, namely Recency, Frequency and Monetary (RFM), the validation results of the k-means algorithm using the Silhouette Coefficient produce a value of 0.937 while the Davies Bouldin Index produces a value of 0.158 and the validation results of the Fuzzy C-Means algorithm using the Silhouette Coefficient produces a value of 0.921 while davies bouldin index returns the value 0.145.

Table 6. The results of the validation of the two algorithms

Algorithm	K	Silhouette Coefficient	Davies Bouldin Index
K-Means	7	0.93	0.15
Fuzzy C-Means	7	0.92	0.14

From the following Table 6, it can be seen that the K-Means algorithm gets the highest validation results using the Silhouette Coefficient, because the K-Means algorithm determines clusters based on the distance and value of data objects in the cluster, where the K-Means characteristics have a relationship with the Silhouette Index calculation method that evaluates the placement of each object.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v3i4.98

in each cluster by comparing the average distance of objects in one cluster and the distance between objects in different clusters.

While the Fuzzy C-Means algorithm gets the highest results using the Davies Bouldin index. because the Fuzzy C-Means algorithm determines clusters based on membership levels between 0 and 1, which causes grouping of data that not only has the same value in a cluster, but also groups of values that have two or more groups according to their membership level, where the characteristics Fuzzy C-Means has a relationship with the Davies Bouldin Index calculation method which maximizes the inter-cluster distance and at the same time tries to minimize the distance between points in a cluster.

Both algorithms produce good clustering and have a small number of validation differences, so it can be said that the clustering process is going well and placing each customer in the right cluster.

CONCLUSIONS AND SUGGESTIONS

Conclusion

From the research that has been successfully carried out by analyzing, implementing, visualizing and validating the results of RFM, K-Means, Fuzzy C-Means, Silhouette Coefficient and Davies Bouldin Index, 7 clusters have been produced which have the best validation values from both algorithms. Cluster 0 is a bronze customer, which is a customer who is not loyal. Cluster 1 is a silver customer, which is a customer who is less loyal. Cluster 2 is a gold customer, which is an ordinary customer. Cluster 3 is a platinum customer, which is a customer who is quite loyal. Cluster 4 is a diamond customer, which is a loyal customer. Cluster 5 is a super customer, which is a customer with high loyalty and cluster 6 is a superstar customer, namely a customer whose loyalty is very high.

The customer loyalty cluster category that has been obtained from the two algorithms shows that with the RFM method it can be seen that the value of customer loyalty to Brand Limback is still low. For this reason, a better management strategy is needed to increase customer loyalty.

Suggestion

This research still has shortcomings that can be improved in further research. Based on the research results and conclusions, the authors provide suggestions for further research. Customer segmentation in this study uses a clustering algorithm that is often used, namely the K-means and Fuzzy C-Means algorithms. Future research can experiment using various other clustering algorithms.

Further research can be developed using the best and updated information by increasing the amount of data using the latest data and using other data mining methods to conduct cross-/up-selling campaigns and market basket analysis. Cross-/up-selling campaigns are used to sell additional or alternative products that are more profitable for customers. Market basket analysis is used to identify related products that customers usually buy together.

REFERENCES

A. Luthfi, N. Anggraini, A. S. (2020). *Statistik E-Commerce 2020* (E. S. L. Anggraini, S. Utoyo, ed.). ©Badan Pusat Statistik/BPS-Statistics Indonesia.

Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering. *Jurnal Terapan Teknologi Informasi*, *2*(1), 23–32. https://doi.org/10.21460/jutei.2018.21.76

Ashari, B. S., Otniel, S. C., & Rianto. (2019). Perbandingan Kinerja K-Means Dengan DSCAN Untuk Metode Clustering Data Penjualan Online Retail. *Jurnal Siliwangi*, *5*(2), 72–77.

Christina Deni Rumiarti, I. B. (2017). Pelanggan Pada Customer Relationship Management Di Perusahaan Ritel: Studi Kasus Pt Gramedia Asri Media. *Jurnal Sistem Informasi (Journal of Information System)*, *v*(Syariah Economic, Zakat), 1–7.

Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2019). Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, *1*(September), 918. https://doi.org/10.30645/senaris.v1i0.100

Kiat, A. B. H., Azhar, Y., & Rahmayanti, V. (2020). Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency, & Monetary). *Jurnal Repositor*, *2*(7), 945–952. https://doi.org/10.22219/repositor.v2i7.973

Nahdliyah, M. A., Widiharih, T., & Prahutama, A. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index. *Jurnal Gaussian*, *8*(2), 161–170.

Nurmalasari, Mukhayaroh, A., Marlina, S., Hartini, S., Muryani, S., Sinnun, A., … Adiwihardja, C. (2020). Implementation of Clustering Algorithm Method for Customer

Segmentation. *Journal of Computational and Theoretical Nanoscience*, *17*(2), 1388–1395. https://doi.org/10.1166/jctn.2020.8815

Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, *1*(1), 72–77. https://doi.org/10.31539/intecoms.v1i1.141

Rustiyan, R., & Mustakim, M. (2018). Penerapan Algoritma Fuzzy C Means untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(2), 171. https://doi.org/10.25126/jtiik.201852605

Sadewo, M. G., Windarto, A. P., & Wanto, A. (2018). Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *2*(No.1 Oktober 2018), 311–319. https://doi.org/10.30865/komik.v2i1.943

Saifullah, S., Zarlis, M., Zakaria, Z., & Sembiring, R. W. (2017). Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *1*(2), 180. https://doi.org/10.30645/j-sakti.v1i2.41

Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokkan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika "MANTIK,"* *4*(1), 22–31. https://doi.org/10.15642/mantik.2018.4.1.22-31

Sutresno, S. A., Iriani, A., & Sediyono, E. (2018). Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan. *Jurnal Teknik Informatika Dan Sistem Informasi*, *4*, 433.

Syahdan, S. Al, & Sindar, A. (2018). Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota. *Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota*, *1*.

Taqwim, W. A., Setiawan, N. Y., & Bachtiar, F. A. (2019). *Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt . Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering*. *3*(2), 1986–1993.

Tempola, F., Muhammad, M., & Mubarak, A. (2020). Penggunaan Internet Dikalangan Siswa SD di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering dan Validasi DBI. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(6), 1153. https://doi.org/10.25126/jtiik.2020722370

Triyansyah, D., & Fitrianah, D. (2018). Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *Jurnal Telekomunikasi Dan Komputer*, *8*(3), 163. https://doi.org/10.22441/incomtech.v8i3.4174

Winarta, A., & Kurniawan, W. J. (2021). Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python. *JTIK (Jurnal Teknik Informatika …*, *5*(1).

Wiranda, L., & Sadikin, M. (2019). Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika*, *8*(3), 184–196.

Yunita, Y., Herman, S., Takwim, A., & Widianto, S. R. (2019). *Independent Research Final Report: A Study Of Comparing Conceptual And Performance Of K-Means And Fuzzy C Means Algorithms (Clustering Method Of Data Mining) Of Consumer Segmentation*. Bandung.