Customer Segmentation Using the RFMD Model and Fuzzy C-Means Algorithm

Muhammad Hafis Zikri
Information System
Faculty of Science and Technology
UIN Sultan Syarif Kasim
Riau
Pekanbaru, Indonesia
12050313115@students.uin-suska.ac.id

Siti Monalisa
Information System
Faculty of Science and Technology
UIN Sultan Syarif Kasim
Riau
Pekanbaru, Indonesia
siti.monalisa@uin-suska.ac.id

Fitriani Muttakin
Information System
Faculty of Science and Technology
UIN Sultan Syarif Kasim
Riau
Pekanbaru, Indonesia
fitrianimuttakin@uin-suska.ac.id

Abstract— Many businesses face challenges in optimizing customer data processing, which often limits the ability to understand customer behavior and improve marketing strategies. This research addresses these challenges by applying the RFMD (Recency, Frequency, Monetary, Diversity) model combined with the Fuzzy C-Means (FCM) clustering algorithm to segment customers based on transaction data. The results identified five distinct customer segments based on Customer portfolio Analysis (CPA), which were validated using the Davies-Bouldin Index (DBI), with each segment showing diverse levels of engagement and behavioral patterns. The results show that the best clusters of Superstar and Golden customers are clusters 4 and 2, while Typical and Occasional customers are clusters 1 and 3. The lowest cluster of Everyday customers is found in cluster 5. The findings provide applicable insights to improve customer retention and optimize data-driven marketing strategies.

Keywords— Clustering, Customer Segmentation, Fuzzy C-Means, RFMD

I. INTRODUCTION

In the information age, product- or customer-based marketing strategies are essential for business competition [1][2]. Segmentation is necessary to group customers with similar characteristics to understand their behavior [3]. As a result, customers become the primary focus of the competition between companies [4][5]. Segmentation groups customers and offers product diversity to maximize revenue [6]. By understanding diverse customer preferences, segmentation becomes an effective approach to managing a company's customers [7].

In the competitive retail industry, businesses often face challenges in effectively leveraging customer transactional data [8]. While such data is typically used for financial reporting, it is rarely utilized to gain a thorough understanding of customer behavior. Therefore, this research focuses on customer retention by performing efficient segmentation.

Juniarti (2021) studied customer segmentation in the company by dividing them into potential customers and loyal customers based on RFM. However, this strategy could be more optimal because customer needs and activities vary. Therefore, segmentation must be optimized by considering recency, frequency, monetary, and diversity [6]. The RFM model is a traditional model marketassesses customer value based on three variables: recency, frequency, and monetary from the customer's perspective [10]. The RFMD model considers product diversity, allowing for better estimation of customer segmentation [11]. This research combines these dimensions using data from one retail company to create a segmentation approach that can be widely adopted by various industries, regardless of sector or company specifics.

This research applies the Fuzzy C-Means (FCM) algorithm as a clustering method to segment customers. The clustering method optimizes the centroid without having identical groups based on sample characteristics [12]. FCM uses a Fuzzy clustering model, allowing data to become members of all classes with membership degrees between 0 and 1 [13] and determining a particular group's membership degree [14]. Research [15] shows that FCM is more effective than K-Means and more suitable for clustering data. Research [16] also confirmed that FCM is more effective based on cluster validity as measured by the Davies Bouldin Index (DBI). Therefore, this study utilizes DBI to assess cluster validity. DBI evaluates clustering success and improves performance in determining the optimal number of clusters [17][18]. The lowest DBI value indicates the most optimal cluster result [19].

Furthermore, customer segmentation analysis uses the Customer Portfolio Analysis (CPA) method to identify high-value customers, which helps make targeted decisions [20]. In this study, 6 clusters were generated. Cluster 5, with the superstar segment category, shows that this cluster has a customer segment with a high loyalty level.

II. METHODOLOGY

This research includes several stages, from data collection, preprocessing, and application of the Fuzzy C-Means algorithm to analysis and interpretation of results. The data used is sales transaction data from January to December 2023 at Printing X in Excel format. The initial stage involves data preprocessing, namely performing data selection and cleaning to resolve and remove missing values. Then, RFMD transformation and data normalization were performed. Next, the Fuzzy C-Means algorithm was used to cluster the data, with evaluation using the Davies-Bouldin Index. The final stage includes ranking the cluster results based on segmentation characteristics and recommending strategies based on the clusters formed. The research methodology is illustrated in Figure 1.

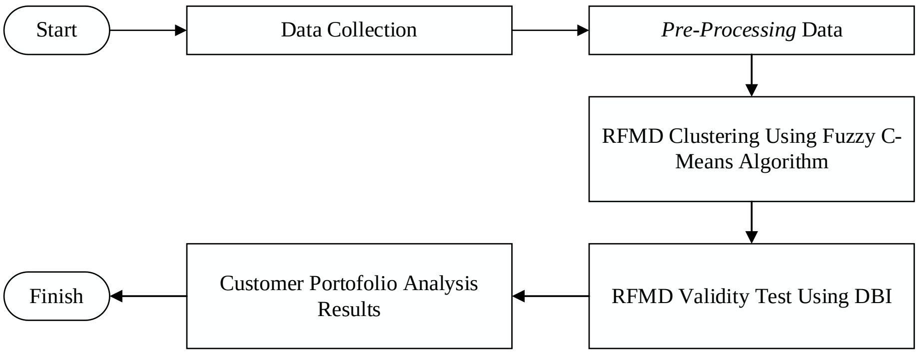

Figure 1 Research Methodology

A. Customer Segmentation

Customer segmentation is a fundamental concept in marketing that aims to differentiate customers, market based on various criteria, and group them retrieved from similar characteristics. The research objectives are to understand and optimize the value or potential of each customer [21]. Customer segment characteristics can be divided into six categories [16]:

1) Superstar Customer: A very high level of loyalty, large transaction (monetary) value, high transaction frequency, and large total transaction value.
2) Golden Customer: The second highest transaction value, high transaction frequency, and good transaction value quality.
3) Typical Customer: Overall average transaction value and value.
4) Occasional Customer: The second lowest transaction value after Dormant Customer, low recency value, but with high transaction frequency.
5) Everyday Shopper: An increasing number of transactions, relatively low transaction value, and transaction value ranging from medium to low.

6) Dormant Customer: The lowest transaction frequency, transaction value, and low recency value.

B. Clustering

Clustering is organizes data into groups, or clusters, that consist of similar items while being different from those in other clusters [22]. This method optimizes the cluster center position based on the similarity of sample characteristics [12]. The main goal is to improve the function between classes and reduce cluster similarity [23]. Clustering is useful for identifying patterns in data and can be applied to classification, image processing, and pattern recognition [24].

C. Fuzzy C Means Algorithm

FCM was initially proposed by Dunn in 1973 and later refined by Bezdek in 1981. FCM is an unsupervised classification algorithm that uses Fuzzy logic [25]. The principle of FCM is to set the center for each cluster, with each data point having a degree of membership to each cluster valued between 0 and 1 [26]. One of the main advantages of fuzzy clustering is its ability to group irregularly distributed objects. The following are the steps of the Fuzzy C-Means algorithm [12]:

1) Input data to be clustered as an n x m matrix.
2) Set the number of clusters (c), partition level (ω), iteration limit (maxIter), minimum error limit (ε), initial value of objective function (Pθ=0), and initial iteration (t=1).
3) Generate a number for the initial partition matrix U.
4) Calculate the center of each k-cluster.
5) Calculate the objective function at the t iteration.
6) Calculate the change in the partition matrix.

D. Recency, Frequency, Monetary, and Diversity Model

The RFM model is a quantitative analysis tool in customer relationship management, developed by Arthur Hughes in 1994 to segment customers based on Recency, Frequency, and Monetary [25]. These three variables include Recency or the time since the last transaction, Frequency or the number of transactions, and Monetary or the total value of customer transactions [26]. The RFMD model adds a Diversity parameter that counts the number of different products purchased, helping to recognize customers interested in new products and increase profitability through diverse offerings [6].

E. Davies-Bouldin Index (DBI)

David L. Davies and Donald W. Bouldin introduced the DBI method in 1979 as an evaluation tool to assess the performance of clustering algorithms. DBI serves as a method to evaluate how effectively clustering is done, considering the number of clusters and the characteristics of the features in the dataset [17]. The main objective of DBI calculation is to maximize the inter-cluster distance [27]. The DBI value is calculated using the following equation [28]:

$$DBI = \frac{1}{k}\sum_{I=1}^{k} \max i \neq j(R\ i, j) \qquad (1)$$

Where k is the number of clusters used.

III. RESULT AND DISCUSSION

A. Data Collection

Secondary data collection is carried out to obtain raw data that will be processed in the segmentation and analysis process. The data obtained from Printing X consists of Microsoft Excel files covering transactions from January to December 2023, with 2,348 rows of data to be processed. The attributes include customer name, transaction date, total shopping, and product category.

B. Pre Processing Data

Before the data is further processed, several steps must be taken to ensure optimal data capture, remove redundancies, ensure completeness, and addressing inconsistencies in the data. In this study, preprocessing involves data selection based on RFMD criteria.

C. Cleaning Data

The cleaning data process, also called data cleaning, is done to reduce noise that can affect the calculation results. The final results of data cleaning are presented in Table 1.

Table 1 Cleaning Data

No	Tanggal	Nama	Total	Kategori
1	23-Oct-2023	X1	Rp,1,200,000	medali
2	27-Sep-2023	X2	Rp,1,200,000	spanduk
3	2-Jan-2023	X3	Rp,70,000	spanduk
4	17-Jul-2023	X4	Rp,50,000	roll up banner
5	15-Nov-2023	X5	Rp,25,000	spanduk
…	…	…	…	…
1021	14-Oct-2023	X1021	Rp325,000	goody bag
1022	20-Jul-2023	X1022	Rp261,000	spanduk
1023	31-Mar-2023	X1023	Rp403,250	spanduk

Based on Table 1, 1023 data can be processed from the original 2,348 rows after data cleaning.

D. Data Transformation

At this stage, customer transaction data is transformed into RFMD criteria based on the results of data cleaning. The RFMD criteria table is presented in Table 2.

Table 2 Data Transformation

No	Nama	R	F	M	D
1	X1	70	1	Rp1,200,000	1
2	X2	96	1	Rp1,200,000	1
3	X3	364	1	Rp70,000	1
4	X4	168	1	Rp50,000	1
5	X5	47	1	Rp25,000	1
…	…			…	…
1021	X1021	79	1	Rp325,000	1
1022	X1022	165	3	Rp261,000	1
1023	X1023	276	3	Rp403,250	2

Table 2 shows the results of the RFMD criteria transformation for normalization in the next calculation after data transformation.

E. Data Normalization

Data normalization aims to reduce the important differences among the variables R, F, M, and D. This is achieved by transforming the variable values into a numerical scale ranging from 0 to 1. The min-max normalization technique can be employed for this purpose. Based on the established RFMD criteria, the minimum and maximum values are presented in Table 3.

Table 3 Min and Max Normalization

	R	F	M	D
MIN	2	1	Rp,1,000	1
MAX	364	107	Rp,106,193,648	8

Normalization is performed for all variables once the minimum and maximum values are known. The variable R (recency) is adjusted with other variables because it has opposite values, other variables show better results with increased value, while the variable R (recency) shows better results with decreased value. Therefore, the normalized recency value is subtracted from 1. The RFMD normalization results are presented in Table 4.

Table 4 Normalization Result

No	Nama	1-NR	NF	NM	ND
1	X1	0.812	0.000	0.011	0.000
2	X2	0.740	0.000	0.011	0.000
3	X3	0.000	0.000	0.001	0.000
4	X4	0.541	0.000	0.000	0.000
5	X5	0.876	0.000	0.000	0.000
…	…			…	…
1021	X1021	0.787	0.000	0.003	0.000
1022	X1022	0.550	0.019	0.002	0.000
1023	X1023	0.243	0.019	0.004	0.143

Table 4 shows the results after data normalization, where all data has been converted into numerical values with a range between 0 and 1 to avoid too large a difference in range.

F. Clustering Using Fuzzy Cmeans Algorithm

The Fuzzy C-Means algorithm is a clustering technique that is widely used in various applications. In this study, the clustering process was conducted using Google Colab, which provides researchers with data processing facilities. Google Colab has multiple features, including libraries such as Pandas and NumPy, and a cmeans function that allows centroid determination for between 3 to 6 clusters. This clustering process was limited to a maximum of 6 clusters, according to the customer segmentation characteristics under study. Therefore, the clustering starts from 3 clusters to 6 clusters. The centroids for each cluster are presented in Table 5, Table 6, Table 7, and Table 8.

Table 5. Centroid for 3 Clusters

Cluster	1-NR	NF	NM	ND
1	0.860952	0.014797	0.013567	0.06881
2	0.167819	0.003682	0.005469	0.01757
3	0.560463	0.007922	0.007551	0.044462

Table 6. Centroid for 4 Clusters

Cluster	1-NR	NF	NM	ND
1	0.853098	0.06033	0.034891	0.367055
2	0.535107	0.005687	0.005976	0.032532
3	0.848062	0.007559	0.010157	0.024769
4	0.157904	0.002975	0.005089	0.014561

Table 7. Centroid for 5 Clusters

Cluster	1-NR	NF	NM	ND
1	0.896825	0.008557	0.010558	0.023288
2	0.448507	0.006045	0.005871	0.036213
3	0.68463	0.005518	0.007152	0.029631
4	0.142407	0.002503	0.005115	0.011074
5	0.858747	0.059296	0.033554	0.375859

Table 8.Centroid for 6 Clusters

Cluster	1-NR	NF	NM	ND
1	0.830692	0.034553	0.016028	0.274221
2	0.127876	0.001615	0.004499	0.002898
3	0.760877	0.003118	0.005189	0.00769
4	0.393272	0.003231	0.004022	0.010984
5	0.580492	0.003021	0.004251	0.009009
6	0.918266	0.004411	0.004966	0.006923

G. Validity Test of Clustering Results

After the clustering process from 3 to 6 clusters is complete, the next step is to use the Davies-Bouldin Index (DBI) approach to validate the clustering results. In the DBI calculation, the data used is the result of the final clustering process. A smaller DBI value, or one close to zero, indicates that the clustering scheme is optimal. The DBI validation values for each cluster are presented in Table 9.

Table 9 DBI Score

No	Cluster	DBI Score
1	3	0.794
2	4	0.778
3	5	0.741
4	6	0.808

Based on the results of the clustering validity test using DBI, the optimal number of clusters to group customer data is five, because it has the smallest DBI value or close to zero. To find out the number of customers from each cluster that has been processed on Google Colab, can be seen in Figure 2.

Figure 2 Number of Members per Cluster

Based on Figure 2, it can be seen the number of customers after clustering using the Fuzzy C-Means algorithm based on the DBI value which produces cluster 1 with 253 customers, cluster 2 with 201 customers, cluster 3 with 271 customers, cluster 4 with 224 customers, and cluster 5 with 74 customers.

H. Analysis of Customer Segment Characteristics

After obtaining the best clusters based on the DBI results, the next step is to examine each group to identify the customer variables associated with each cluster after achievingby ch optimal results. This result data is then used to describe the characteristics of the customers in each cluster. The characteristics of customer segments can be identified by analyzing the purchasing patterns of customers within each cluster based on their RFMD quality. The range of RFMD variable values within each group is then determined through analyzing the data within each cluster. The MIN() and MAX() functions in Microsoft Excel were used to find the range of values. The range of each cluster is presented in Table 10 below.

Table 10. Range of RFMD Variable Values

Cluster	R	F	M	D
1	2-77	1-31	10000-25000000	1-2
2	159-243	1-8	11000-14900000	1-4
3	79-159	1-15	3500-19950000	1-2
4	259-364	1-6	1000-7500000	1-4
5	3-149	3-107	70000-106200000	2-8

The characteristics of each customer segment will be formed based on the analysis of purchasing patterns within each group, using RFMD attributes as a guide. The following are the results of the thatsis that has been carried out:

1) Superstar Customer is located in cluster 4 with a total of 224 customers. Superstar customers are customers with very high value and strong loyalty. Their Recency (R) is high, indicating they have recently made a transaction. Although their shopping Frequency (F) is low, they tend to make high-value purchases in the low to medium Monetary (M) category each time they transact.. They have high Diversity (D), which indicates interest in a wide range of products or services. These customers are a very valuable segment for companies and have great potential to try new products, so they are suitable for VIP programs or special rewards.
2) Golden Customer is located in cluster 2 with a total of 201 customers. Golden customers are high-value customers with fairly good engagement. They have a high Recency (R) value, meaning they have recently made a transaction. Their Frequency (F) is low to medium, indicating they do not transact frequently but tend to be consistent. In terms of Monetary (M), their spending value is medium to high, meaning that they make sizable purchases. Their Diversity (D) is quite varied, meaning they are interested in a wide range of products or services. Customers in this category are loyal customers who contribute significant revenue and can be considered for upgrading to “Superstar” customers.
3) Typical Customer is located in cluster 1 with a total of 253 customers. This category has quite varied shopping patterns. In terms of Recency (R), they are moderate, meaning they make transactions not too long ago but not too often. Their transaction Frequency (F) varies, indicating their inconsistent shopping intensity.. In terms of Monetary (M), the total value of their purchases can range from low to high, depending on the specific need or occasion. For Diversity (D), the product categories they buy are usually limited. These customers are the average type and may make purchases only when there is a promotion or a specific need.
4) Occasional Customer is located in cluster 3 with a total of 271 customers. This category includes customers who make occasional purchases. They have a medium Recency (R) value, meaning their last transaction is neither too close nor too far away. Their transaction Frequency (F) is low to medium, indicating they shop infrequently. Their Monetary (M) is in the low to medium range, so the total value of their spending

is not very large. The Diversity (D) of products or services they purchase is also limited. These customers are likely to be swayed by promotions or discounts and could be targets for retention programs aimed at increasing transaction frequency or value.

5) Everyday Customer is located in cluster 5 with a total of 74 customers. Everyday customers are very active customers with regular shopping patterns. Their Recency (R) value is low, indicating they make frequent transactions. Their transaction Frequency (F) is very high, which means they are regular customers with repeat transactions. Their Monetary (M) is also very high, indicating they shop in large quantities. Their Diversity (D) is medium, which means they buy from several product categories but could be more diverse. These customers are highly loyal and a stable source of revenue for the company, making them suitable for loyalty programs that reward continued engagement.

IV. CONCLUSION

Based on the research that has been done, it is obtained that 5 clusters are the best number of clusters based on DBI cluster validation with a value of 0.741, Customer Portfolio Analysis (CPA) including customers belonging to the Superstar and Golden segments, namely clusters 4 and 2 with high potential that deserve more attention through award programs to maintain and increase loyalty. Meanwhile, the Typical and Occasional segments, namely clusters 1 and 3, require the right promotional strategy to increase transaction frequency and value. Everyday customers, cluster 5, with high transaction frequency and large shopping value, are an important foundation for the company's revenue and mustinconsistenttent be maintained through sustainable loyalty programs. The findings provide strategic guidance for companies in crafting better marketing approaches.

REFERENCES

[1] A. A. D. Sulistyawati and M. Sadikin, “Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan,” *Sistemasi*, vol. 10, no. 3, p. 516, 2021, doi: 10.32520/stmsi.v10i3.1332.

[2] G. V. Bhupathiraju, V. Reddygari, and T. Sai, “Data mining for the online retail industry : Customer segmentation and assessment of customers using RFM and k-means ABSTRACT : INTRODUCTION : RESEARCH OBJECTIVES :,” pp. 1–9, 2022.

[3] W. V. Oliveira, D. S. A. Araujo, and L. C. T. Bezerra, “Supermarket customer segmentation: A case study in a large Brazilian retail chain,” *Proc. - 2022 IEEE 24th Conf. Bus. Informatics, CBI 2022*, vol. 1, pp. 70–79, 2022, doi: 10.1109/CBI54897.2022.00015.

[4] D. P. Hidayatullah, R. I. Rokhmawati, and ..., “Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada …,” *... Inf. dan Ilmu ...*, vol. 2, no. 8, 2018.

[5] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, “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, doi: 10.30880/ijie.2019.11.03.018.

[6] M. Y. Smaili and H. Hachimi, “New RFM-D classification model for improving customer analysis and response prediction,” *Ain Shams Eng. J.*, vol. 14, no. 12, p. 102254, 2023, doi: 10.1016/j.asej.2023.102254.

[7] Y. O. Pailan, Y. H. Chrisnanto, and A. I. Hadianna, “Segmentasi Loyalitas

Pelanggan Berbasis RFM (Recency, Frequency, Monetary) Menggunakan K-Means pada PD. Persada Ikan," *Pros. Semin. Nas. Sist. Inf. dan Teknol.*, vol. 4, no. 1, pp. 167–171, 2020.

[8] Y. Huang, M. Zhang, and Y. He, "Research on improved RFM customer segmentation model based on K-Means algorithm," *Proc. - 2020 5th Int. Conf. Comput. Intell. Appl. ICCIA 2020*, pp. 24–27, 2020, doi: 10.1109/ICCIA49625.2020.00012.

[9] Y. Juniarti, "ANALISIS SEGMENTASI PELANGGAN BERDASARKAN RECENCY, FREQUENCY, MONETARY (RFM) DAN DEMOGRAFI MENGGUNAKAN ALGORITMA DBSCAN," Skripsi tidak ditampilkan, Program Studi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim, RIAU, 2021.

[10] A. Dwirahmawan, N. Y. Setiawan, and D. Pramono, "Analisis Metode Fuzzy C-Means dengan Model LRFM terhadap Segmentasi Pelanggan pada Toko Butik (Studi Kasus : Diva Collection Kota Cilegon)," vol. 5, no. 12, pp. 5278–5287, 2021.

[11] R. Heldt, C. S. Silveira, and F. B. Luce, "Predicting customer value per product: From RFM to RFM/P," *J. Bus. Res.*, vol. 127, no. xxxx, pp. 444–453, 2021, doi: 10.1016/j.jbusres.2019.05.001.

[12] F. Juniati, S. Monalisa, R. Zafa, and M. Muslim, "KLASTERISASI CUSTOMER LIFETIME VALUE DENGAN MODEL LRFM MENGGUNAKAN ALGORITMA Fuzzy C-Means," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 6, no. 1, p. 38, 2020, doi: 10.24014/rmsi.v6i1.8635.

[13] A. Ilham, N. Y. Setiawan, and T. Afirianto, "Analisis Segmentasi Pelanggan Kartu Prabayar Kabupaten Malang dengan RFM Model Menggunakan Metode Fuzzy C-Means Clustering (Studi Kasus : PT . XYZ)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 8, pp. 2487–2498, 2020.

[14] S. S. Prasetyo, A. R. Hakim, D. Statistika, F. Sains, and U. Diponegoro, "PENERAPAN FUZZY C-MEANS KLUSTER UNTUK SEGMENTASI PELANGGAN E-COMMERCE DENGAN METODE RECENCY FREQUENCY MONETARY (RFM)," *J. GAUSSIAN*, vol. 9, pp. 421–433, 2020.

[15] K. Zhou and S. Yang, "Effect of cluster size distribution on clustering: a comparative study of k-means and fuzzy c-means clustering," *Pattern Anal. Appl.*, vol. 23, pp. 455–466, 2020.

[16] D. L. Aditya and D. Fitrianah, "Comparative Study of Fuzzy C-Means and K-Means Algorithm for Grouping Customer Potential in Brand Limback," *J. Ris. Inform.*, vol. 3, no. 4, pp. 327–334, 2021, doi: 10.34288/jri.v3i4.241.

[17] M. N. Dayat, N. Suarna, and Y. A. Wijaya, "Analisa Clustering untuk Mengelompokan Data Penayangan Film Bioskop Menggunakan Algoritma K-Means," *Intern. (Information Syst. Journal)*, vol. 6, no. 1, pp. 68–78, 2023, doi: 10.32627/internal.v6i1.686.

[18] W. Gie and D. Jollyta, "Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, no. April, pp. 187–191, 2020.

[19] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.

[20] E. Ditendra, S. Monalisa, S. Anderjovi, and S. Lesmana, "KLASTERISASI CLV DENGAN MODEL LRFM MENGGUNAKAN ALGORITMA FUZZY C-MEANS (Studi Kasus: Pangeran Gym Pekanbaru)," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 6, no. 1, p. 109, 2020, doi: 10.24014/rmsi.v6i1.8535.

[21] 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.*, vol. 33, no. 10, pp. 1251–1257, 2021, doi: 10.1016/j.jksuci.2018.09.004.

[22] A. Nofiar, S. Defit, and Sumijan, "Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering," *J. KomtekInfo*, vol. 5, no. 3, pp. 1–9, 2019, doi: 10.35134/komtekinfo.v5i3.26.

[23] R. Muliono and Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *J. Comput. Eng. Syst. Sci.*, vol. 4, no. 2, pp. 2502–714, 2019.

[24] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.

[25] Y. Huang, "Research on improved RFM customer segmentation model based on K-Means algorithm," *Int. Conf. Comput. Intell. Appl.*, no. August 2020, pp. 24–27, 2020, doi: 10.1109/ICCIA49625.2020.00012.

[26] F. Carneiro, R. M. A. Moreira, V. N. De Gaia, R. R. Frias, and V. Miguéis, "Applying Data Mining Techniques and Analytic Hierarchy Process to the Food Industry : Estimating Customer Lifetime Value," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, no. Kotler 2000, pp. 266–277, 2021.

[27] F. Tempola and A. F. Assagaf, "Clustering of Potency of Shrimp In Indonesia With K-Means Algorithm And Validation of Davies-Bouldin Index," vol. 1, no. Icst, pp. 730–733, 2018, doi: 10.2991/icst-18.2018.148.

[28] R. K. Dinata, H. Novriando, N. Hasdyna, and S. Retno, "Reduksi Atribut Menggunakan Information Gain untuk Optimasi Cluster Algoritma K-Means," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 1, p. 48, 2020, doi: 10.26418/jp.v6i1.37606.