Customer Segmentation Analysis Through RFM-D Model and K-Means Algorithm

Refri Martiansah
Information System
Faculty of Science and Technology
UIN Sultan Syarif Kasim Riau
Pekanbaru, Indonesia
12050313004@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

Mona Fronita
Information System
Faculty of Science and Technology
UIN Sultan Syarif Kasim Riau
Pekanbaru, Indonesia
monafronita@uin-suska.ac.id

Abstract— This research analyzes customer segmentation through the RFM-D (Recency, Frequency, Monetary, and Diversity) model and the K-Means algorithm. The data comes from sales transactions at Café Z from January 2023 to February 2024, with 10,212 entries. The applied methodology includes several stages: data pre-processing, cleaning, transformation, normalization, and clustering. Clustering validation was carried out using the Davies-Bouldin Index (DBI) to ensure the quality of the clusters formed. The analysis results identified three customer clusters based on purchasing behavior, indicating that the K-Means algorithm effectively groups customers. These findings provide insight for companies to design marketing strategies that are more focused and appropriate to the characteristics of each customer segment. Companies can improve operational efficiency, increase customer satisfaction, and maximize profitability by utilizing this segmentation. This research contributes to optimizing resource allocation and personalizing marketing approaches, ultimately strengthening customer relationships.

Keywords—RFM-D, customer segmentation, K-Means, purchasing behavior analysis, marketing strategy.

I. INTRODUCTION

The rapid development of information technology has brought significant changes in various aspects of life, including the world of marketing[1]. Customers play a crucial role in business strategy as individuals or groups who purchase products or services based on their own decisions, consider benefits and prices, and serve as the primary source of revenue for the company[2]. Therefore, efficient management of sales data becomes essential to support operational decisions and more effective marketing strategies[3].

Many companies prioritize a personalized approach to foster close relationships with customers, such as greeting them by name and remembering their preferences[4]. However, this approach still has weaknesses because it is not supported by in-depth data analysis of customer behavior, such as visit frequency, purchase amounts, and the diversity of products bought[5]. Without adequate analysis, companies miss the opportunity to identify purchasing patterns and design more targeted marketing strategies[6].

Customer segmentation analysis through the RFM-D model (Recency, Frequency, Monetary, and Diversity) can provide deeper insights into customer characteristics and behaviors[7]. Companies can design more efficient and targeted marketing strategies by identifying customer segments based on visit frequency, purchase value, and product diversity [8]. This approach also allows companies to optimize resources and enhance profitability[9].

To address the shortcomings of existing personal approaches, customer segmentation analysis using data mining techniques, such as clustering with the K-Means algorithm, can be employed to group customers based on specific characteristics[10]. This clustering can help companies identify more specific customer segments and develop marketing strategies that are more aligned with the characteristics of each segment[11]. By utilizing this method, companies can enhance operational effectiveness and efficiency while maximizing profits[12].

Based on previous research, customer segmentation using the RFM and RFM-D models has proven effective in enhancing marketing strategies[3]. This study applies the K-Means algorithm for clustering using the RFM and RFM-D models to analyze customer segmentation in a company[13]. The results of this analysis are expected to help the company develop more targeted marketing strategies, enhance customer satisfaction, and maximize profitability through a more data-driven approach and more accurate segmentation[8].

II. METHODOLOGY

This research involves several structured stages, starting with data collection, including sales transaction data from Cafe Z in Excel from January 2023 to February 2024. A data preprocessing stage involves data selection and cleaning to address missing values. Next, RFM-D clustering is performed using the K-Means algorithm to group customer data. The validity testing of RFM-D is carried out using the Davies-Bouldin Index to evaluate the clustering results. Finally, customer segmentation analysis is performed based on the formed clusters, and strategy recommendations are developed according to the characteristics of each cluster. The methodology process is illustrated in Figure 1.

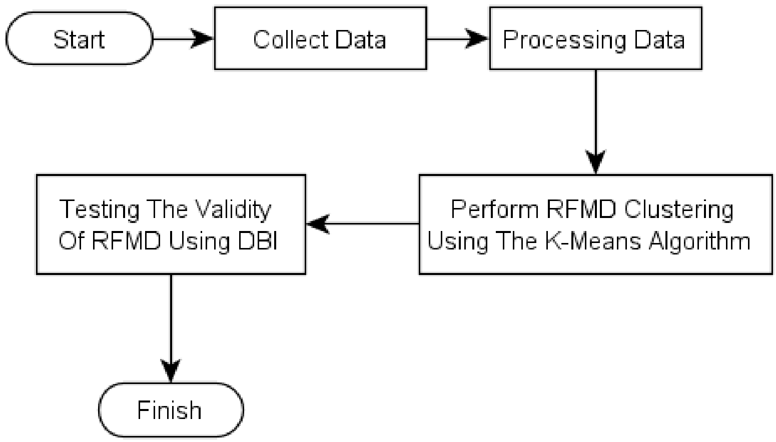

Figure 1 Research Methodology

A. Customer Segmentation

Customer segmentation is the grouping of customers based on several aspects, such as customer details, historical data, customer activity, customer interests, and customer satisfaction. This information is essential for better understanding the value of customers and gaining insights into customer behavior[14]. Customer segmentation is conducted to identify profitable customers for the company[15].

Six customer characteristics are selected and arranged based on the concepts explained in the article.[1]:

1) Loyal Customers: Recency (R): High value; recently purchased. Frequency (F): High value; frequently makes purchases. Monetary (M): High value; spends significant money. Diversity (D): High value; purchases a variety of products.
2) Potential Customers: Recency (R): High value; recently purchased. Frequency (F): Low value; infrequently makes purchases. Monetary (M): Low value; spends a small amount of money. Diversity (D): High value; open to trying new products.
3) One-Time Customers: Recency (R): High value; recently purchased. Frequency (F): Low value; only made one purchase. Monetary (M): Varies; spending can be low or high. Diversity (D): Low value; purchases a limited variety of products.
4) Inactive Customers: Recency (R): Low value; haven't purchased in a long time. Frequency (F): Varies; may have shopped frequently but now rarely does. Monetary (M): Varies; previous spending might have been high but is now low. Diversity (D): Varies; may have purchased a variety of products in the past but does so no longer.
5) At-Risk Customers: Recency (R): Low value; haven't purchased in a long time. Frequency (F): Low value; infrequently makes purchases. Monetary (M): Low value; spends a small amount of money. Diversity (D): Low value; purchases a limited variety of products.
6) New Customers: Recency (R): High value; recently purchased. Frequency (F): Low value; hasn't made many purchases yet. Monetary (M): Varies; initial spending can be low or high. Diversity (D): Varies; may have tried a few new products.

B. Data Mining

Data Mining (DM) is the process of exploring and analyzing data automatically or semi-automatically to discover meaningful patterns and rules[16]. This technique helps reveal the structure of datasets, such as clustering customers with similar needs and behaviors[17]. It is often referred to as Knowledge Discovery in Databases (KDD) [18], Data Mining is used to explore large datasets and extract useful hidden information, which is crucial for making significant business decisions[19]. This process aims to excavate and analyze data to uncover the patterns of information contained within it[20].

C. Clustering

Clustering is a technique in data analysis used to group a set of objects or data into clusters based on similarities or specific characteristics[15]. The goal of clustering is to identify patterns within the data, facilitate understanding, and assist in decision-making, such as in marketing or predicting customer behavior[21]. In this method, we use the distance between points or between clusters, such as:

- Single Linkage :

$$d(C_i, C_j) = \min\big(d(x, y)\big), \quad x \in C_i,\ y \in C_j$$

- Complete Linkage :

$$d(C_i, C_j) = \max(d(x, y))$$

- Average Linkage :

$$d\left(C_i, C_j\right) = \frac{1}{|C_i||C_j|1} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y)$$

To evaluate the clustering results, we can use the Silhouette Score:

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

Where:

- $a(i)$ = average distance between point i to all points in the same cluster.
- $b(i)$ = the average distance between point i to all points in the nearest cluster.

Value of S(i):

- Closer to 1 → Good clustering.
- Closer to 0 → Point is on the border of two clusters.
- Closer to -1 → Bad clustering.

D. Model RFM-D.

Based on the journal[1] , The RFM-D model improves the traditional RFM model used for customer analysis and segmentation. RFM itself represents three primary parameters:

Recency (R): Measures how recently a customer has purchased. The more recent the purchase, the higher the recency score. This helps identify active customers who are likely to make repeat purchases.

$$R = T - t_i$$

Where:

- T = analysis date (current time).
- t_i = customer i's last transaction date.
- The smaller the R value, the more active the customer is.

Frequency (F): Measures how often a customer purchases within a specific period. Customers who shop frequently receive a higher score, indicating their loyalty to the brand or product.

$$F = Number\ of\ customer\ transactions\ in\ a\ certain\ period$$

- The higher the F, the more loyal the customer.

Monetary (M): This measure measures a customer's total spending over a certain period. Customers who spend more money receive a higher score, indicating their value to the company.

$$M = \sum_{j=1}^{n} p_j$$

- p_j = the jth transaction value of the customer.
- The higher M is, the more valuable the customer is to the business.

Addition of the Diversity (D) Parameter: This parameter reflects the variety of product types customers purchase. Customers who buy a diverse range of products will receive a higher score for diversity, indicating they are more open to trying new products.

$$D = t_{last} - t_{first}$$

- t_{last} = the customer's last transaction date.
- t_{first} = the date of the customer's first transaction.
- The larger D is, the longer the customer stays in business.

E. K-Means

K-means is a clustering algorithm first introduced by J. MacQueen in 1967. Since then, it has become one of the most popular methods for data analysis and grouping[22]. K-Means is a simple and easy-to-understand clustering algorithm that is fast and efficient in processing large datasets[23]. The main formula used in K-Means is to calculate the distance between data points and the cluster center (centroid) using Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

The steps of the K-Means algorithm are as follows[23]:

1. Initialization: Determine the desired number of clusters (k). Randomly select (k) data points from the dataset as the initial centroids.
2. Cluster Assignment: For each data point in the dataset, calculate the distance (usually using Euclidean distance) from that data point to each centroid. Assign the data point to the cluster with the nearest centroid.
3. New Centroid Calculation: After all data points are assigned to clusters, calculate the new centroid for each cluster by taking the average of all data points assigned to that cluster.
4. Convergence Condition: Check if the new centroids are the same as the previous centroids (or if the change in centroids is below a certain threshold). If yes, the algorithm has converged and can be stopped; if not, return to step 2.
5. Output: Once convergence is achieved, the final output assigns each cluster's centroids and data points to their respective clusters.

F. Davies Bouldin Index (DBI)

The validity testing method uses the Davies-Bouldin Index, which calculates the average value of each point in the dataset. Knowing the optimal cluster results with the smallest DBI value or close to a non-negative value (≥ 0) indicates that the clusters are improving. To find the Davies-Bouldin Index (DBI), use the following equation[24].

$$D\ B\ I = \frac{1}{N} \sum_{i=1}^{N} \max_{j \neq i} R_{ij}$$

Where:

- N = Number of clusters
- R_{ij} = Similarity between clusters i and j, which is calculated by:

$$R_{ij} = \frac{S_i + S_j}{d(C_i, C_j)}$$

  - S_i = Average distance between each point in cluster i and the centroid of cluster i (also called dispersion or spread of the cluster).

$$S_i = \frac{1}{[C_i]} \sum_{x \in C_i} d(x, C_i)$$

  - $d(C_i, C_j)$ = Distance between cluster i centroid and cluster j centroid (e.g. using Euclidean Distance).

$$d(C_i, C_j) = \|C_i - C_j\|$$

  - $\max(R_{ij})$ → For each cluster i, we find the largest R_{ij} of all clusters j ≠ i.

III. RESULT AND DISCUSSION

A. Data Collection

The raw data obtained from Café Z in this study consists of 10,212 rows of records in Excel format, covering the period from January 2023 to February 2024. The attributes processed include date, item, price, and customer name. This data encompasses transactions relevant to customer segmentation analysis.

B. Pre Processing data

The preprocessing stage involves removing irrelevant, incomplete, and duplicate data. This process ensures the data is clean and relevant, leading to more accurate analysis results. After this stage, the data is ready for further analysis.

C. Cleaning Data

The data cleaning process ensures that the data used in the analysis is relevant and accurate. The final results of data cleaning are shown in Table 1.

Table 1 Cleaning Data

No	Date	Item	Price	Customer Name
1	01-01-2023	FRIED RICE-CHICKEN	22.000	Z1
2	01-01-2023	INDOMIE-DOUBLE	18.000	Z2
3	01-01-2023	BURGER/KEBAB-BEFF	15.000	Z3
4	01-01-2023	COFFEE-COFFEE MILK PALM SUGAR	20.000	Z4
5	01-01-2023	RICE BOWL-GOAT	30.000	Z5
…	…	…	…	…
…	…	…	…	…
10211	10-02-2024	FRIED RICE-CHICKEN	22.000	Z10211
10212	10-02-2024	FRIED RICE-GOAT	30.000	Z10212

Table 1 shows that after the cleaning process, the amount of data remained consistent, indicating that there were no significant issues with the data collected for the transformation stage. This process ensures that the data used to calculate attributes such as Recency, Frequency, Monetary, and Diversity (RFM-D) meets cleanliness standards and is ready for further processing in the transformation stage.

D. Data Transformation

Based on Table 1, Data was transformed to change the data from Table 1 into the RFM-D criteria format. In this table, each customer is represented by a unique name, without duplication, along with the relevant metrics: Recency (R), Frequency (F), Monetary (M), and Diversity (D). The results of RFM-D data transformation are presented in Table 2.

Table 2 Data Transformation

No	Customer Name	R	F	M	D
1	Z1	314	262	5.312.000	62
2	Z2	361	252	5.049.000	56
3	Z3	321	150	3.582.000	47
4	Z4	319	77	1.334.000	29
5	Z5	321	160	3.446.000	53
…	…	…	…	…	…
…	…	…	…	…	…
220	Z220	313	3	40.000	3
221	Z221	313	3	65.000	2

Table 2 is a strong foundation for normalization process described in Table 3 and Table 4. This Table shows the variations in customers' R, F, M, and D values . The values in this Table make it possible to identify each attribute's minimum and maximum values.

E. Data Normalization

The next stage is normalization, as shown in Table 2. Based on this table, a normalization process is carried out to reduce the scale differences between the R, F, and M attributes, making them more comparable. The range used is 0-1, achieved by applying the min-max normalization formula. The results of normalizing all the data are presented in Table 3. This normalization is crucial to ensure that all attributes have a balanced influence in the clustering analysis.

Table 3 Min Max Normalization

	R	F	M	D
MIN	313	2	15000	2
MAX	713	353	7630000	67

After the minimum and maximum values are determined, the R variable is adjusted to reflect the values of 1−R. This adjustment is necessary because other variables indicate better results with larger values, while the R variable indicates better results with smaller values. The results of this adjustment can be seen in Table 4.

Table 4 Normalization 1-R

No	Customer Name	1-NR	F	M	D
1	Z1	0.9975	0.7407	0.6956	0.9231
2	Z2	0.8800	0.7123	0.6611	0.8308
3	Z3	0.9800	0.4217	0.4684	0.6923
4	Z4	0.9850	0.2137	0.1732	0.4154
5	Z5	0.9800	0.4501	0.4506	0.7846
…	…	…	…	…	…
…	…	…	…	…	…
220	Z220	1.0000	0.0028	0.0033	0.0154
221	Z221	1.0000	0.0028	0.0066	0.0000

Table 4 displays data that is ready for clustering analysis in the next process. All values have been adjusted to a varying range from 0 to 1 to make grouping customers based on their purchasing behavior easier.

F. Clustering Using the K-Means Algorithm

After the data is normalized, the next stage is to implement the K-Means algorithm using Python. By grouping customer transaction data using the K-MeansS_i algorithm for each cluster as the average distance between points in the cluster and its centroid. Next, calculate the distance between cluster centroids and similarityR_{ij} for each pair of clusters. DBI is calculated based on the average maximum similarity value for each cluster, with the smallest DBI value (0.562) occurring in the three clusters which show the most optimal configuration. To improve DBI results, it is necessary to optimize the number of clusters and perform better data preprocessing by removing outliers and normalizing features. Additionally, testing various distance metrics and performing relevant feature selection can help improve the clustering structure. The distribution of customers within each cluster is detailed, providing valuable insight into purchasing patterns. This step is carried out to identify optimal clusters based on the Davies-Bouldin Index (DBI) of the clusters formed.

Table 5 DBI Score

Cluster	DBI Value
2	0.639
3	0.562
4	0.653
5	0.728
6	0.726
7	0.679
8	0.750
9	0.727
10	0.720

Table 5 shows the best number of clusters that will be used to carry out the clustering process, which can be seen in Table 6.

Tabel 6 Best cluster average value

Cluster	Cluster Member	R	F	M	D
0	68	0.3525	0.0510	0.0522	0.1760
1	107	0.8707	0.0321	0.0336	0.1082
2	46	0.9665	0.4548	0.4451	0.6471

Table 6 shows the average values of the attributes R (Recency), F (Frequency), M (Monetary), and D (Diversity) for each cluster formed after applying the K-Means algorithm. Based on the calculated average value of customers in each cluster, this table provides an initial overview of its characteristics for customer segmentation.

G. Customer Segmentation Analysis

After obtaining the clustering results in Table 6, the next step is to examine customer purchasing patterns in each cluster by returning the RFM-D values to their original form before normalizing, to find out the range of RFM-D values in each cluster. The RFM-D and K-Means methods offer comprehensive behavioral analysis and efficiency in data grouping, compared to simpler demographic approaches, more in-depth but difficult to access psychographic segmentation, and hierarchical clustering which are slower and more complex in interpretation. Table 7 shows the distance between RFM-D values for each cluster.

Table 7 Customer Segmentation

No	Cluster	R	F	M	D
1	Cluster 0	468 – 713	3 – 96	45000-201000	3 – 36
2	Cluster 1	313 – 470	2 – 77	15000 - 1334000	2 - 29
3	Cluster 2	313 – 387	77 – 353	1482000 - 7630000	27 - 67

After determining the range of RFM-D values, customer segment characteristics will be formed by analyzing purchasing behavior in each cluster using the RFM-D attributes. The results of the analysis are presented in Table 8.

Table 8 Purchase Behavior Analysis

No	Cluster	Information
1	0	Customers in this cluster have a high R-value and low frequency, indicating that they are customers who have not transacted for a long time but have the potential to return.
2	1	Customers in this cluster have a higher purchase frequency but lower monetary value, indicating they may need incentives to increase their spending.
3	2	This cluster of highly active customers indicates loyalty and potential for further purchases. Marketing strategies can be focused on enhancing their engagement even more.

IV. CONCLUSION

The results of the segmentation analysis indicate that the RFM-D approach and K-Means algorithm effectively identify customer purchasing patterns, offering a comprehensive analysis of customer behavior and efficient data grouping compared to simpler demographic approaches and more complex hierarchical clustering methods. Recommendations for the company include:

1. Implement targeted marketing strategies based on the characteristics of each cluster.
2. Develop loyalty programs for customers in Cluster 2 to maximize profitability.
3. Utilize data analysis to enhance the customer experience and increase purchase frequency in Cluster 1.

REFERENCES

[1] 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.

[2] D. R. Ramadhan and N. Rokhman, "Segmentation-Based Sequential Rules For Product Promotion Recommendations As Sales Strategy (Case Study: Dayra Store)," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 3, p. 243, 2020, doi: 10.22146/ijccs.58107.

[3] A. Wibowo and A. R. Handoko, "Segmentasi Pelanggan Ritel Produk Farmasi Obat Menggunakan Metode Data Mining Klasterisasi Dengan Analisis Recency Frequency Monetary (RFM) Termodifikasi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, p. 573, 2020, doi: 10.25126/jtiik.2020702925.

[4] M. Zanker, L. Rook, and D. Jannach, "Measuring the impact of online personalisation: Past, present and future," *Int. J. Hum. Comput. Stud.*, vol. 131, no. January, pp. 160–168, 2019, doi: 10.1016/j.ijhcs.2019.06.006.

[5] T. Aldyah, A. Sugiono, M. Muzakkar, and J. Heikal, "Rfm Segmentation Analysis For Determine Online Marketing Strategy:The Soul Coffee Mate Case Study," *Budg. J. Business, Manag. Account.*, vol. 5, no. 2, pp. 797–807, 2024.

[6] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 204–215, 2021, doi: 10.24002/konstelasi.v1i1.4293.

[7] V. Dawane, P. Waghodekar, and J. Pagare, "RFM Analysis Using K-Means Clustering to Improve Revenue and Customer Retention," *SSRN Electron. J.*, no. Icsmdi, 2021, doi: 10.2139/ssrn.3852887.

[8] Y. Wicaksono, "Segmentasi Pelanggan Bisnis Dengan Multi Kriteria Menggunakan K-Means," *Indones. J. Bus. Intell.*, vol. 1, no. 2, p. 45, 2019, doi:

10.21927/ijubi.v1i2.872.
[9] M. A. Satriawan, R. Andreswari, and O. N. Pratiwi, "Segmentasi Pelanggan Telkomsel Menggunakan Metode Clustering Dengan Rfm Model Dan Algoritma K-Means Telkomsel Customer Segmentation Using Clustering Method With Rfm Model and K-Means Algorithm," *e-Proceeding Eng.* , vol. 8, no. 2, pp. 2876–2883, 2021.
[10] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
[11] R. Y. Firmansah, J. Dedy Irawan, and N. Vendyansyah, "Analisis Rfm (Recency, Frequency and Monetary) Produk Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 334–341, 2021, doi: 10.36040/jati.v5i1.3282.
[12] M. Fithriyah, M. A. Yaqin, and S. Zaman, "K-Means Clustering Untuk Segmentasi Produk Berdasarkan Analisis Recency, Frequency, Monetary (RFM) Pada Data Transaksi Penjualan," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 3, no. 2, pp. 151–164, 2021, doi: 10.28926/ilkomnika.v3i2.284.
[13] S. I. Murpratiwi, I. G. Agung Indrawan, and A. Aranta, "Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail," *J. Pendidik. Teknol. dan Kejuru.*, vol. 18, no. 2, p. 152, 2021, doi: 10.23887/jptk-undiksha.v18i2.37426.
[14] A. Z. Putri, M. Afdal, S. Monalisa, and I. Permana, "Penerapan Algoritma Fuzzy C-Means Pada Segmentasi Pelanggan B2B dengan Model LRFM," *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1423–1432, 2023, doi: 10.30865/mib.v7i3.6150.
[15] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 32, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.
[16] K. L. Tsui, V. Chen, W. Jiang, F. Yang, and C. Kan, "Data Mining Methods and Applications," *Springer Handbooks*, pp. 797–816, 2023, doi: 10.1007/978-1-4471-7503-2_38.
[17] A. Novikov, "PyClustering: Data Mining Library," *J. Open Source Softw.*, vol. 4, no. 36, p. 1230, 2019, doi: 10.21105/joss.01230.
[18] Annisa Nadaa Shabrina, M. Afdal, and Siti Monalisa, "Comparison Of K-Means, K-Medoids, and Fuzzy C-Means Algorithms for Clustering Drug User's Addiction Levels," *J. Sist. Cerdas*, vol. 6, no. 2, pp. 113–122, 2023, doi: 10.37396/jsc.v6i2.313.
[19] I. Ahmad, S. Samsugi, and Y. Irawan, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/107
[20] P. Palupiningsih and B. Prayitno, "Implementasi Algoritma Fp-Growth Untuk Penentuan Rekomendasi Produk UMKM berdasarkan Frekuensi Pembelian," *J. Teknoinfo*, vol. 17, pp. 493–501, 2023.
[21] M. R. A. Purnomo, A. R. Anugerah, A. Azzam, A. U. Khasanah, and M. N. Alfareza, "RFM-based customers clustering for precise industrial marketing strategy formulation," *J. Adv. Manuf. Technol.*, vol. 15, no. 1, pp. 1–14, 2021.
[22] J. Wu *et al.*, "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K -Means Algorithm," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8884227.
[23] M. Aliyev, E. Ahmadov, H. Gadirli, A. Mammadova, and E. Alasgarov, "Segmenting Bank Customers via RFM Model and Unsupervised Machine Learning," 2020, [Online]. Available: http://arxiv.org/abs/2008.08662

[24] R. Siagian, P. Sirait, and A. Halim, “The Implementation of K-Means dan K-Medoids Algorithm for Customer Segmentation on E-commerce Data Transactions,” *Sistemasi*, vol. 11, no. 2, p. 260, 2022, doi: 10.32520/stmsi.v11i2.1337.