INFOMATEK: Jurnal Informatika, Manajemen dan Teknologi
Volume 28 Nomor 1, Juni 2026

APPLICATION OF THE RFM MODEL AND K-MEANS CLUSTERING FOR CUSTOMER SEGMENTATION IN E-WALLET TOP-UP SERVICES

Agus Sundari[1]*, Indra Syah Putra[2], Nuraini Sibuea[3]

[1]Computer Science, Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia
[2]Information Technology, Universitas Wirahusada Medan, Indonesia
[3]Mathematics, Universitas Riau, Indonesia

Abstract: The implementation of digital payment technology through e-wallet top-up services requires financial institutions to understand user characteristics and behavior comprehensively The objective of this study is to segment customers based on their e-wallet top-up behavior by analyzing 143,836 bill payment transaction records using the RFM (Recency, Frequency, Monetary) model combined with the K-Means clustering algorithm. The dataset contains more than one hundred thousand transaction entries, with RFM parameters representing the time since the last transaction, the frequency of top-ups, and the monetary value spent by users. The RFM scoring process is applied to quantify user activity levels before entering the clustering stage. The K-Means clustering model successfully grouped customers into three distinct segments. The first segment represents low-activity users, the second consists of moderately active customers with stable transaction behavior, while the third segment captures highly engaged users with the highest transaction frequency and value. Evaluation metrics, including a silhouette score of 0.64, a Calinski-Harabasz index of 21690.50, and a Davies-Bouldin score of 0.70, demonstrate strong clustering performance and reliable separation between groups. The findings provide valuable insights for designing service strategies, improving mobile banking system performance, and developing targeted marketing approaches tailored to each customer segment. This research highlights the potential of RFM based clustering as a decision-support tool for enhancing digital payment service optimization and customer engagement.

Keywords: Customer segmentation, Recency, Frequency, Monetary, K-Means Clustering

I. INTRODUCTION

In the banking sector, customer retention and satisfaction are vital to long term business sustainability. One effective approach to bolstering loyalty and enhancing organizational performance is through customer segmentation (Sundari et al., 2024). As the financial industry becomes increasingly digital, institutions must develop a deeper, more personalized understanding of customer preferences and behaviors. This perspective is supported by scholars who argue that segmentation fundamentally aims to rationally identify and address customers needs and desires (Das & Nayak, 2022).

Insights into the traits associated with customer loyalty are also essential for elevating service quality (Puspitasari et al., 2020). Among the most prevalent analytical tools in this domain is the RFM (Recency, Frequency, Monetary) model, widely adopted across industrie including retail, healthcare, and banking (Firdaus & Utama, 2021). RFM allows businesses to classify customers according to their strategic value (Anitha &

*) agussundari@uinbukittinggi.ac.id

Received: 10 February 2026
Revised: 28 April 2026
Accepted:28 April 2026
DOI: 10.23969/infomatek.v28i1.42246

Patil, 2022), making behavioral modeling especially critical in banking to drive service improvements (Egorova et al., 2022).

The rapid rise of digital finance in Indonesia evidenced by the surging adoption and transaction volume of e-wallet has introduced new complexities. Despite this growth, banks still struggle to comprehensively understand e-wallet users profiles and underlying needs. This gap underscores the necessity of advanced segmentation strategies that go beyond basic categorization to uncover meaningful behavioral patterns, enabling more precise and effective service delivery.

To achieve such precision, machine learning has emerged as a powerful enabler. Capable of processing vast datasets, uncovering latent patterns, and forecasting future behaviors, machine learning is already applied across financial contexts from fraud detection and credit risk assessment (Machado & Karray, 2022) to portfolio optimization (Al Janabi, 2022), customer segmentation (Das & Nayak, 2022), and churn prediction (Haddadi et al., 2022).

A particularly popular unsupervised machine learning technique for segmentation is K-Means clustering (Lestari et al, 2022). It is frequently integrated with the RFM framework to group customers based on similarities in transactional behavior, such as purchase recency, frequency, and monetary value. This synergy allows banks to discern distinct segment characteristics, interpret behavioral trends, and tailor strategies to convert casual users into loyal customers.

Building on this foundation, the current study utilizes anonymized general customer data to implement an RFM based segmentation model, followed by K-Means clustering analysis (Syahputra et al, 2020). The goal is to generate more accurate, actionable customer segments that offer strategic insights for enhancing service quality and strengthening customer retention in the banking industry.

This study aims to segment customers using the Recency, Frequency, Monetary (RFM) model combined with the K-Means clustering algorithm. By utilizing e-wallet top-up transaction data, this study seeks to identify distinct customer groups based on their historical transaction behavior.

II. RESEARCH METHOD

This study employs a quantitative approach by integrating the Recency, Frequency, Monetary (RFM) model with the K-Means Clustering algorithm to segment customers based on e-wallet top-up transactions. The dataset, obtained internally from a regional bank, consists of 143,836 transactions collected over a two-month period, with customer identities handled in accordance with data privacy standards. It includes attributes such as transaction time, transaction value, and e-wallet type. The analysis involves RFM feature extraction, normalization, clustering, and interpretation, where Recency measures days since the last transaction, Frequency indicates the number of transactions, and Monetary reflects total transaction value. All analyses are conducted using Python and its analytical libraries. K-Means is used to group customers with similar behavior, with the optimal number of clusters determined through evaluation techniques, aiming to generate meaningful segments for improving service delivery and customer retention strategies.

III. RESULT AND DISCUSSION

E-wallet top-up customers can come from various segments, including individuals, MSMEs, and corporations. Each group has different needs and behaviors in using e-wallet services. Customer segmentation for e-wallet top-ups using the K-Means clustering

algorithm can be performed based on e-wallet transaction data, such as total transaction value, transaction type, transaction date, and transaction frequency. The K-Means clustering algorithm will classify customers into several segments based on their transaction patterns, with each segment having its own unique characteristics.

By thoroughly understanding the data, researchers can produce more accurate and meaningful customer segmentation, which can assist banks in improving their e-wallet top-up services. The available dataset is in CSV format, containing 143,836 rows and 7 columns. The features contained in the dataset are as follows:

1. **Date Time** : the transaction date, datetime type.
2. **Biller Number** : the customer's payment type code, numeric type.
3. **Transaction Value** : the amount of the transaction made by the customer, numeric type.
4. **Host Ref** : the transaction reference code, object type.

The initial bill payment dataset can be seen in figure 1.

	Date Time	Biller Number	Transaction Value	Host Ref	Last Rc
0	2024-02-01 00:00:00	OVO	250000	9.0	0
1	2024-02-01 00:00:00	OVO	100000	40.0	0
2	2024-02-01 00:00:00	OVO	50000	51.0	0
3	2024-02-01 00:02:00	Gopay	50000	171117.0	0
4	2024-02-01 00:03:00	OVO	250000	248.0	0
...	...	...	...	...	...
143831	2024-03-31 23:58:00	Gopay	100000	96117.0	0
143832	2024-03-31 23:58:00	OVO	1000000	91095.0	0
143833	2024-03-31 23:59:00	OVO	250000	91130.0	0
143834	2024-03-31 23:59:00	OVO	100000	91141.0	0
143835	2024-03-31 23:59:00	OVO	100000	NaN	0

143836 rows × 5 columns

Figure 1. Dataset

To show the insights gained from a more detailed data analysis, the author uses the matplotlib module for visualization in the Python software. The insights can be seen below :

a. Pie chart of e-wallet top-up transaction types

In the context of e-wallet top-ups, a pie chart can be used to illustrate the percentage contribution of each type of e-wallet to the total transactions. This helps companies quickly visualize and understand the distribution of e-wallet top-up transactions, providing insights into user preferences for specific payment methods and platforms. A pie chart of e-wallet top-up transaction types can be seen in figure 2.

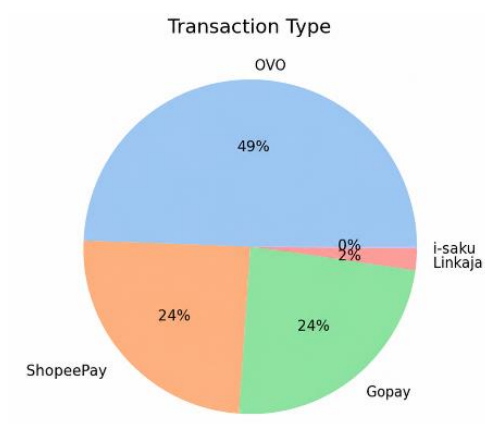

Figure 2. Pie chart of e-wallet

In Figure 2, it can be seen that based on transaction data over a two-month period, most Bank customers used the OVO top-up transaction type (49%), followed by ShopeePay and GoPay transactions, each accounting for 24% of the total transactions. In addition to these dominant e-wallets, the dataset also includes other e-wallet services, namely i-saku and LinkAja, although their usage is significantly lower, with i-saku contributing approximately 2% and LinkAja accounting for a negligible proportion. Overall,

the e-wallets identified in the dataset are OVO, ShopeePay, GoPay, i-saku, and LinkAja, with OVO being the most widely used and LinkAja the least utilized.

b. Bar Chart of e-wallet top-up transaction types

In the context of e-wallet top-up transaction types, a bar chart can be used to show the number or value of transactions made using various types of e-wallets or payment platforms. Each bar will represent the value or amount of a specific transaction category. This diagram provides a clear overview of different transaction preferences and volumes, making it easier to analyze and make decisions regarding marketing strategies or product development for e-wallet top-ups. The bar chart for e-wallet top-up transaction types can be seen in figure 3.

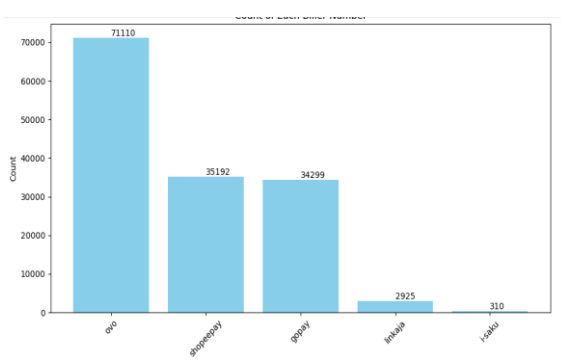

Figure 3. Bar chart of e-wallet

In Figure 3, it can be seen that Bank customers topped up their OVO e-wallets the most, with a total of 71,110. Customers topped up their ShopeePay e-wallets 35,192 times. Customers topped up their GoPay e-wallets 34,299 times. Customers topped up their LinkAja e-wallets 2,925 times. Customers topped up their i-Saku e-wallets 310 times.

c. Bar Chart of E-Wallet Top-Up Transaction Types by Hour

Using this diagram, companies can quickly see the peak times for e-wallet top-up transactions. The bar chart showing the type of transaction for each hour can be seen in Figure 4.

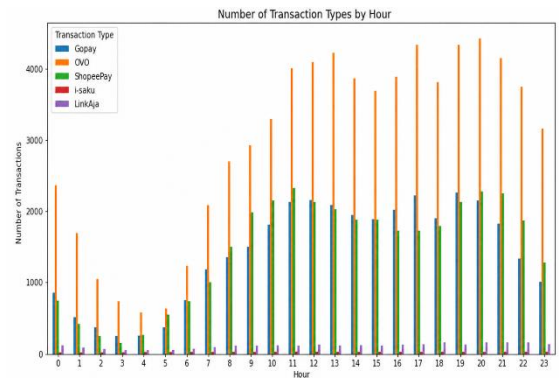

Figure 4. Bar chart e-wallet by hour

In Figure 4, it can be seen that over a 2-month period, the most customers used OVO e-wallet for top-ups compared to other brands, and the most payments were made at 8 PM. The fewest transactions were made at 4 AM.

Modeling is an important stage in the process of segmenting e-wallet top-up customers based on the variables of recency (the time since the customer's last transaction), frequency (the number of transactions made by the customer over a period), and monetary value (the amount of money spent by the customer over a specific time) using the K-Means clustering algorithm. At this stage, the K-Means clustering model is trained to group Bank e-wallet top-up customers into several segments based on their transaction characteristics. The segmentation results can help the bank improve its services and products to meet the diverse needs of its customers.

After exploratory data analysis, the next step is to calculate recency, frequency, and monetary

value. The results obtained can be seen in Table 1.

Table 1. RFM

No.	Recency	Frequency	Monetary
0	39	1	250000
1	4	6	6000000
2	17	1	100000
3	53	1	250000
4	4	11	1210000
…	…	…	…
21189	0	2	500000
21190	9	4	40000
21191	2	11	1100000
21192	28	3	270000
21193	20	1	50000

After performing RFM modeling calculations, the next step is to calculate the RFM Score. The RFM score is an analysis method used to group customers based on three main dimensions: Recency, Frequency, and Monetary value. The RFM score helps companies segment customers, prioritize marketing and service efforts based on customer value and potential, and improve retention and profitability with more focused strategies. As for the RFM Score, it can be seen in Table 2.

No.	R_Score	F_Score	M_Score	RFM_score
0	1	1	2	4
1	4	4	5	13
2	3	1	1	5
3	1	1	2	4
4	4	5	4	13
…	…	…	…	…
21189	5	3	3	11
21190	3	4	1	8
21191	5	5	4	14
21192	2	3	3	8
21193	2	2	1	5

After performing the RFM Score, the next step is to look at the bar charts for each RFM. The RFM bar charts can be seen in figure 5.

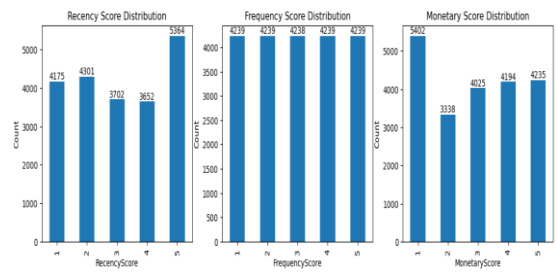

Figure 5. RFM Score

In Figure 5, it can be seen that in the recency section, most customers are scored 5. In the frequency section, most customers are scored 1, 2, 4, and 5. In the monetary section, most customers are scored 1. The labels for each RFM score can be seen in figure 6.

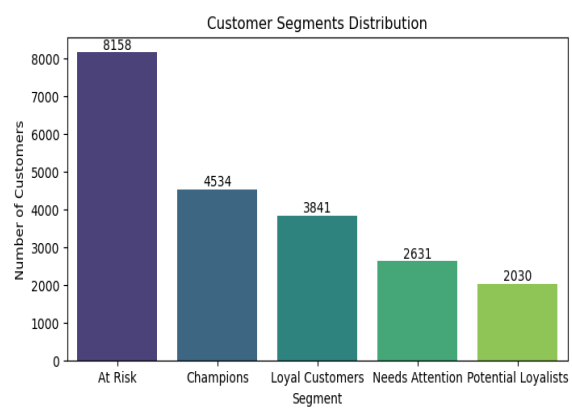

Figure 6. Label RFM Score Segment

After performing the RFM calculations above, the next step is clustering using K-Means

clustering to group the bank's e-wallet top-up customers. From the clustering process, there are 3 customer clusters. As can be seen in Figure 7.

The interpretation of these clustering results in the Figure 7 is as follows: Cluster 0: Customers who rarely top up their e-wallets, also known as Silver customers. These customers have high recency values, with lower average top-up amounts and transaction frequency compared to Platinum and Gold customers, but they still represent a valuable customer base.

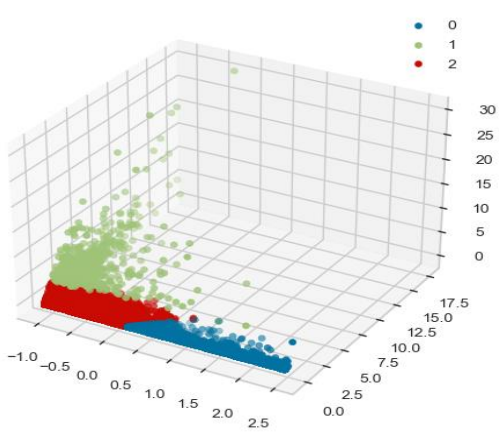

Figure 7. Customer Cluster

Cluster 1: Customers who consistently top up their e-wallets, also known as Gold customers. These customers may have slightly lower average top-up amounts compared to Platinum customers, but they still provide significant revenue for the bank.

Cluster 2: Customers who frequently top up their e-wallets, also known as Platinum customers. This is the bank's most valuable customer, exhibiting the highest transaction frequency, the greatest number of top-ups, a low recency value, and overall engagement with the e-wallet service. They are typically tech-savvy, early adopters of new features, and have high incomes.

Evaluation is an important step in segmenting e-wallet top-up customers using the K-Means clustering algorithm at the Bank. Evaluation helps ensure that the resulting segmentation is accurate, useful, and stable. The evaluation metrics obtained for K-Means clustering are shown in figure 8.

```
For n_clusters=3, the silhouette score is 0.6411227651448275
For n_clusters=3, the calinski score is 21690.50533854461
For n_clusters=3, the davis_bouldin score is 0.70037621882028
```

Figure 8. Evaluation Metrics

Based on the results of the three metrics above, it can be concluded that the K-Means clustering algorithm successfully produced good quality data grouping for e-wallet top-up customer segmentation at the Bank. As can be seen, the number of customers who top up their e-wallets at Bank SUMUT, which has been clustered in Figure 9.

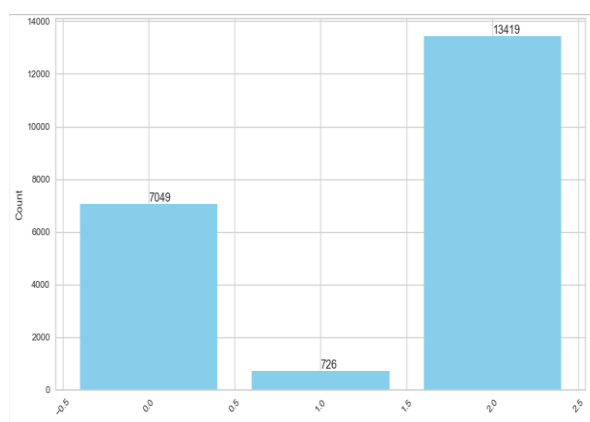

Figure 9. Number of Each Cluster

In Figure 9, it can be seen that there are 7,049 customers in cluster 0, 726 customers in cluster 1, and 13,419 customers in cluster 2.

IV. CONCLUSION

Based on the results of the clustering analysis using the K-Means clustering algorithm, it was concluded that there is a grouping of customers based on the e-wallet top-up transactions made by the customers. The clustering process resulted in 3 clusters: Silver customers, Gold customers, and Platinum customers. The evaluation metric results obtained are: silhouette score: 0.64, Calinski-Harabasz index: 21690.50, and Davis-Bouldin score: 0.70. This result can be used to develop more effective service strategies and improve the customer experience when using the e-wallet top-up payment service within the Bank's M Banking application.

REFERENCES

Al Janabi, M. A. (2022). Optimization algorithms and investment portfolio analytics with machine learning techniques under time-varying liquidity constraints. *Journal of Modelling in Management, 17*(3), 864–895

Anitha, P., & Patil, M. M. (2022). RFM Model for Customer Purchase Behavior Using K-Means Algorithm. *Journal Of King Saud University - Computer And Information Sciences, 34*(5), 1785–1792.

Das, S., & Nayak, J. (2022). Customer Segmentation via Data Mining Techniques: State-of-the-Art Review. *Smart Innovation, Systems and Technologies, 281*(July), 489–507.

Egorova, E., Glukhov, G., & Shikov, E. (2022). Customer Transactional Behaviour Analysis Through Embedding Interpretation. *Procedia Computer Science, 212*(C), 284–294.

Firdaus, U., & Utama, D. N. (2021). Development of Bank's Customer Segmentation Model Based on Rfm+B Approach. *ICIC Express Letters, Part B: Applications, 12*(1), 17–26.

Haddadi, S. J., Mohammadi, M. O., Bahrami, M., Khoeini, E., Beygi, M., & Khoshkar, M. H. (2022). Customer churn prediction in the Iranian banking sector. *The 2022 International conference on applied artificial intelligence* (pp. 1–6). IEEE.

Lestari, D., Fauzan, A. C., & Harliana, H. (2022). Penerapan Algoritma Pillar Untuk Optimasi Penentuan Titik Awal Centroid Pada Algoritma K-Means Clustering. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, *6*(1), 15-24.

Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications, 200*, Article 116889

Puspitasari, N., Widians, J. A., & Setiawan, N. B. (2020). Customer Segmentation Using Bisecting K-Means Algorithm Based on Recency, Frequency, and Monetary (RFM) Model. *Jurnal Teknologi dan Sistem Komputer, 8*(2), 78–83.

Sundari, A., Lydia, M. S., & Muchtar, M. A. (2024, September). Customer Segmentation Based on Recency, Frequency, Monetary, Variety and Duration (RFMVD). In *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)* (pp. 1-5). IEEE.

Syahputra, S., Ramadani, S., & Pardede, A. M. H. (2020). Menentukan Strategi Promosi Menggunakan Algoritma Clustering K-Means. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, *4*(1), 7-14.