

Customer Segmentation of E-Wallet Top-Up Users Based on RFM and K-Means Clustering

Feri Ranja^{1*}, Aser Heber Ginting², Indra Syah Putra³, Roswitha Bukit⁴

^{1,2,3} Faculty of Technology and Business Education, Information Technology and, Universitas Wirahusada Medan, Medan, Indonesia.

⁴ Faculty of Technology and Business Education, Management, Universitas Wirahusada Medan, Medan, Indonesia

Email: ^{1*}feriranja@uwah.ac.id, ²aser.gtg@uwah.ac.id, ³indrasyahputra@uwah.ac.id, ⁴roswithabukit@uwah.ac.id

(*Email corresponding author: feriranja@uwah.ac.id)

Received: February 18, 2026 | Revision: February 20, 2026 | Accepted: February 20, 2026

Abstract

The rapid growth of digital payment systems has significantly increased the use of e-wallet services, creating challenges for banks in understanding customer transaction behavior and developing targeted marketing strategies. This study aims to segment e-wallet top-up customers using the Recency, Frequency, and Monetary (RFM) model integrated with the K-Means clustering algorithm. The research follows the CRISP-DM framework, covering business understanding, data preparation, modeling, and evaluation stages. The dataset consists of 143,836 bill payment transaction records collected from a government bank in North Sumatra over a two-month period. RFM values were calculated to measure customer engagement and transaction value, followed by clustering analysis to group customers based on behavioral similarity. The results identified three distinct customer segments: Silver, Gold, and Platinum. Evaluation metrics indicate that the clustering model produced stable, meaningful segmentation, providing strategic insights to support personalized marketing initiatives and to improve customer retention and service optimization.

Keywords: Customer segmentation, E-wallet, RFM, K-Means, Data mining.

1. INTRODUCTION

The advancement of digital technology has significantly transformed the financial industry, particularly in payment systems. One of the most rapidly growing innovations is the use of digital wallets (e-wallets) as a non-cash transaction instrument. E-wallets facilitate customers in conducting various financial transactions, such as bill payments, purchases, and balance top-ups, efficiently and conveniently. Consequently, banking institutions are encouraged to adapt by providing accessible and seamless e-wallet top-up services.

However, the increasing number of e-wallet users presents new challenges for banks. The competition among financial institutions has become increasingly intense, requiring banks to better understand customer behavior in order to deliver targeted services. One approach to understanding customer behavior is customer segmentation, defined as the process of grouping customers based on similar characteristics or transactional behaviors. Through effective segmentation, banks can design more personalized, efficient, and effective marketing and service strategies. Wedel and [1]) State that market segmentation aims to identify homogeneous customer groups so that companies can develop more targeted and value-based marketing strategies.

In this context, the RFM (Recency, Frequency, Monetary) method is widely used to analyze customer behavior based on three primary dimensions: how recently a customer made a transaction (Recency), how often transactions are made (Frequency), and the total transaction value (Monetary). These three variables provide a comprehensive representation of customer engagement and value. [2] argue that the RFM model is effective for transaction-based customer segmentation and can be integrated with clustering techniques to enhance grouping accuracy. Furthermore, [3]) emphasize that value-based approaches such as RFM support data-driven marketing decisions and assist companies in identifying customers with high customer lifetime value [4].

To improve clustering accuracy, RFM analysis can be combined with the K-Means clustering algorithm, an unsupervised learning technique effective in grouping data into clusters based on attribute similarity. Tsipitsis and [5] explain that integrating RFM with data mining techniques such as K-Means is a common approach in modern Customer Relationship Management (CRM) systems to enhance customer retention and loyalty. By combining RFM and K-Means, this study aims to segment e-wallet top-up customers based on their transaction patterns into categories such as Silver, Gold, and Platinum. The results are expected to support banks in formulating more targeted marketing strategies and improving customer loyalty and satisfaction [6].

2. RESEARCH METHODOLOGY

2.1 Methods And Implementation Stages

The stages of the research implementation are illustrated in the following figure:

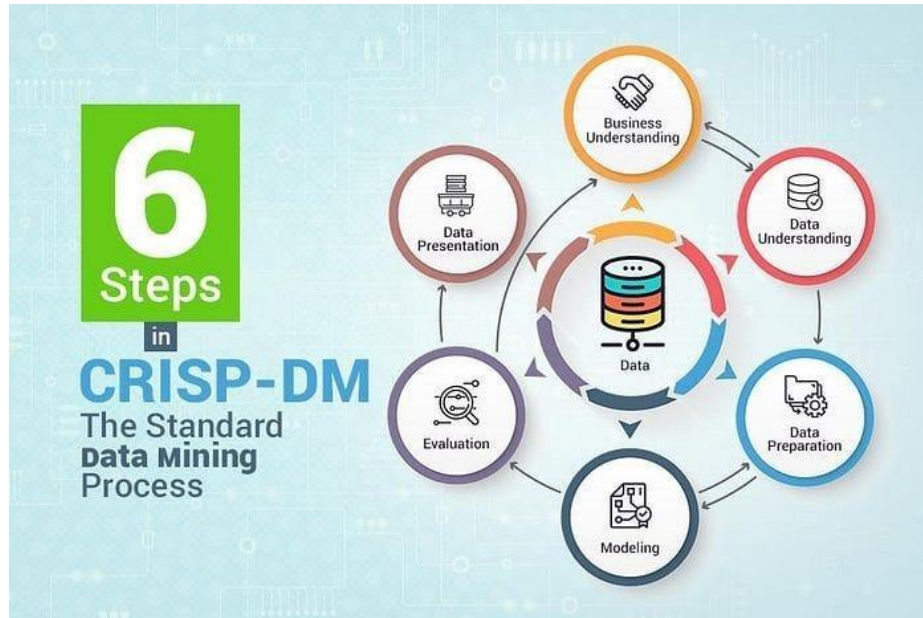


Figure 1. Research Implementation Flow Diagram

The research implementation stages follow the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, which is a standard methodology in data mining projects. CRISP-DM provides a systematic approach consisting of six main phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, ensuring a structured and measurable analytical process[7].

2.1 Business Understanding

The business understanding phase focuses on comprehending the business objectives and defining the problem to be solved. In this study, the problem is how to group e-wallet top-up customers based on their transaction patterns to enable banks to design more effective marketing strategies. [3] explain that customer segmentation aims to identify homogeneous customer groups so that companies can tailor service and promotional strategies more precisely. The e-wallet business is part of the digital transformation in the financial sector, enabling electronic payment transactions. The growing use of e-wallets encourages banks to analyze transaction behavior using data-driven approaches to enhance customer loyalty and value.

2.2 Data Understanding

Data understanding involves analyzing and interpreting data to extract valuable insights. This stage is essential in segmenting e-wallet top-up customers using the K-Means algorithm. A comprehensive understanding of data enables more accurate and meaningful segmentation results. The dataset used in this study is in CSV format, containing 143,836 rows and 8 columns derived from bill payment transactions of a government bank in North Sumatra from February to March 2024. [5] emphasize that accurate transaction data processing significantly determines clustering quality in RFM analysis. Due to privacy concerns, only feature descriptions are presented.

The dataset features include:

1. **Phone Number**, representing the customer's mobile phone number, stored in numeric format.
2. **Date Time**, representing the transaction date and time, stored in date-time format.
3. **Source Account**, representing the customer's bank location code, stored in numeric format.
4. **Biller Number**, representing the payment type code, stored in numeric format.
5. **Customer** representing the destination account number of the customer's transaction, stored in numeric format.
6. **Transaction Value**, representing the transaction amount conducted by the customer, stored in numeric format.
7. **Host Ref**, representing the transaction reference code, stored in object (string) format.

2.3 Data Preparation

In this phase, the database structure is prepared to facilitate the mining process using Jupyter Notebook. The stages include:

1. **Data Selection**
Selecting relevant data and attributes aligned with the mining objectives.
2. **Data Preprocessing**
Ensuring the quality of the data selected during the data selection stage is essential. At this stage, the primary issues addressed include noisy data and missing values. The data cleansing process is conducted using simple query methods to identify and detect potential data anomalies that may still exist within the system.
3. **Data Transformation**
Grouping the selected attributes and calculating the Recency, Frequency, and Monetary values for each customer. [8] explain that transforming transactional data into RFM values enables companies to measure customer value based on transaction history.

2.3 Modeling

Modeling involves segmenting customers based on RFM variables using the K-Means clustering algorithm. K-Means groups data according to proximity to cluster centroids. The integration of RFM and K-Means has been widely applied in customer segmentation research due to its ability to produce objective numerical-based clustering results[9].

3. RESULTS AND DISCUSSION

3.1 RFM Modeling Results

After exploratory data analysis, RFM values were calculated, followed by RFM scoring. RFM scoring groups customers based on recency, frequency, and monetary dimensions to prioritize marketing and service efforts. Rhitungan recency, frequency, and monetary.

Table 1. RFM Calculation Results

	Phone Number	Recency	Frequency	Monetary
0	6,28E+10	39	1	250000
1	6,28E+10	4	6	6000000
2	6,28E+10	17	1	100000
3	6,28E+10	53	1	250000
4	6,28E+10	4	11	1210000
...
21189	6,29E+13	0	2	500000

After completing the RFM modeling process, the next step is to calculate the RFM score. The RFM score is an analytical method used to classify customers based on three primary dimensions: Recency, Frequency, and Monetary value. This scoring approach enables organizations to segment customers, prioritize marketing and service efforts according to customer value and potential, and enhance retention and profitability through more targeted strategies. The RFM score results are presented in Table 2

Table 2. RFM Score

	Phone Number	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_score
0	6,28E+10	39	1	250000	1	1	2	4
1	6,28E+10	4	6	6000000	4	4	5	13
2	6,28E+10	17	1	100000	3	1	1	5
3	6,28E+10	53	1	250000	1	1	2	4
...
21189	6,29E+13	0	2	500000	5	3	3	11

After calculating the RFM scores, the subsequent step is to examine the bar charts for each RFM dimension. The RFM bar charts are presented in Figure 2.

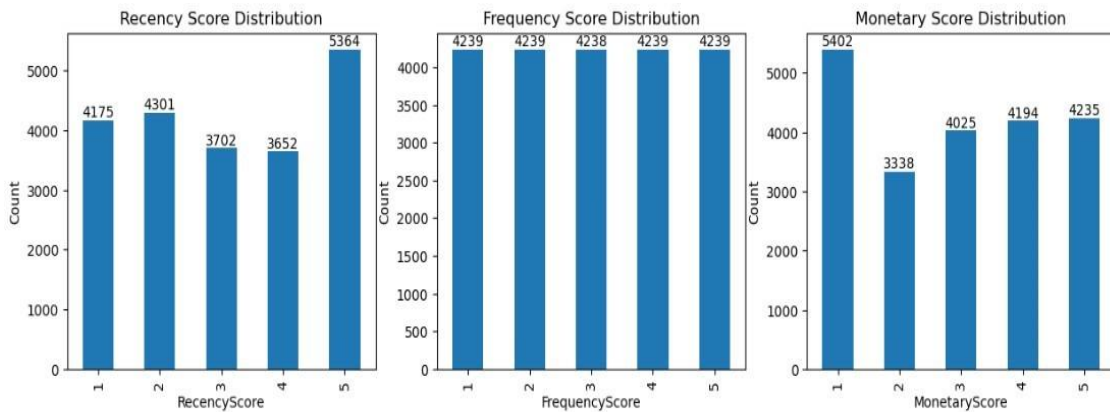


Figure 2. RFM Bar Chart

As illustrated in Figure 2, most customers fall into score 5 in the Recency dimension. In the Frequency dimension, the majority of customers are distributed across scores 1, 2, 4, and 5. Meanwhile, in the Monetary dimension, most customers are concentrated in score 1. The labels for each RFM score category are presented in Figure 3.

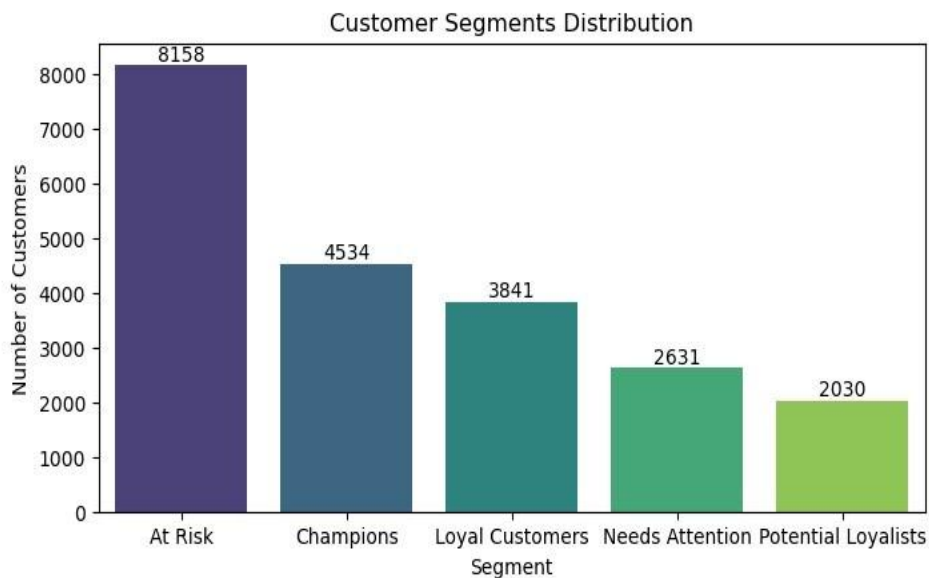


Figure 3. RFM Score Segment Labels

3.2 Clustering Results Using the K-Means Algorithm

After performing the RFM calculations described above, the next step is to conduct clustering using the K-Means algorithm to segment e-wallet top-up customers. The clustering process resulted in three distinct customer clusters, as illustrated in Figure 4

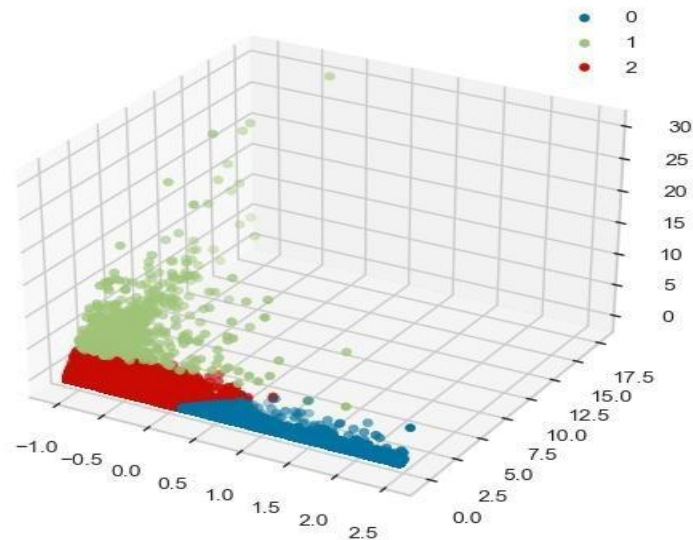


Figure 4. Customer Clusters

The interpretation of the clustering results is as follows:

1. Cluster 0: Customers who rarely perform e-wallet top-up transactions, categorized as *Silver* customers. These customers exhibit high recency values, with lower average top-up amounts and transaction frequencies compared to other categories. Nevertheless, they still represent a valuable customer base. Nasabah yang jarang melakukan transaksi *top-up e-wallet* atau dikategorikan sebagai nasabah *Silver*.
2. Cluster 1: Customers who consistently conduct e-wallet top-up transactions, categorized as *Gold* customers. They tend to have slightly higher average top-up amounts than Silver customers and contribute a steady and significant revenue stream to the bank.
3. Cluster 2: Customers who frequently perform e-wallet top-up transactions, categorized as *Platinum* customers. These represent the most valuable segment for the bank, characterized by the highest transaction frequency, relatively low recency values, and strong overall engagement with digital wallet services. They are typically technologically proficient, early adopters of new features, and possess higher income levels.

3.3 Evaluation

Evaluation is a crucial stage in customer segmentation of e-wallet top-up transactions using the K-Means clustering algorithm in the banking context. The evaluation process ensures that the resulting segmentation is accurate, meaningful, and stable. The evaluation metrics obtained from the K-Means clustering are presented in Figure 5.

```
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 5. Evaluation Metrics

Based on the results of the three evaluation metrics presented above, it can be concluded that the K-Means clustering algorithm successfully produced high-quality data groupings for the segmentation of e-wallet top-up customers in the bank [4].

3.4 Data Presentation

Data presentation is the process of organizing, presenting, and communicating information in data in a visual or verbal manner so it can be more easily understood by readers or specific stakeholders. The primary objective of data presentation is to convey findings or insights effectively, clearly, and in a relevant manner. The number of bank customers who conducted e-wallet top-up transactions and have been clustered is presented in Figure 6.

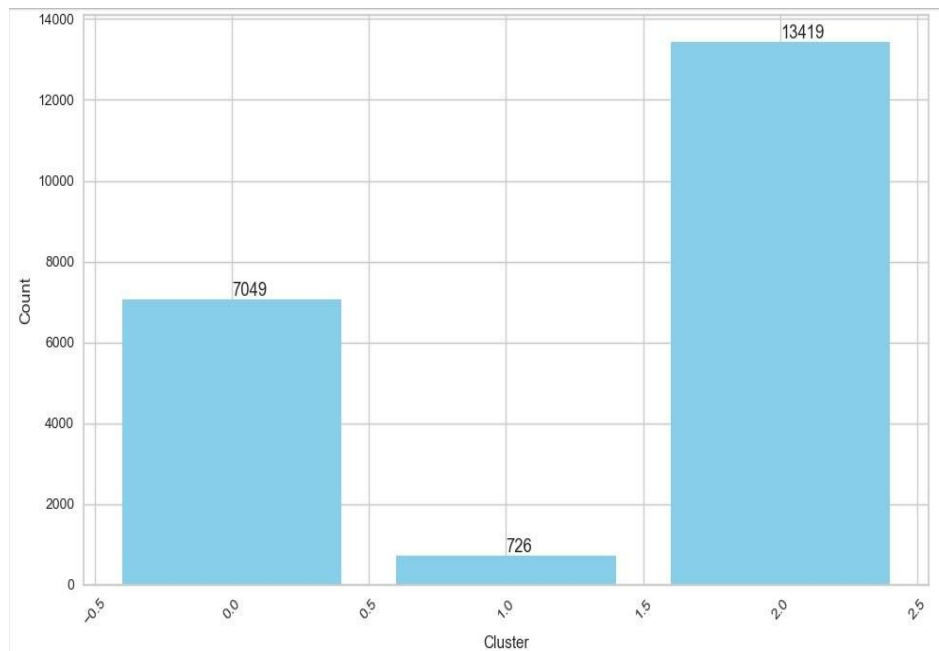


Figure 6. Number of Customers in Each Cluster

As shown in Figure 3.5, Cluster 0 consists of 7,049 customers, Cluster 1 comprises 726 customers, and Cluster 2 includes 13,419 customers.

4. CONCLUSION

The research process followed the CRISP-DM framework, beginning with business understanding and concluding with evaluation. The dataset consisted of bill payment transactions from a government bank in North Sumatra over a two-month period, totaling 143,836 records. The stages conducted included data understanding, data preparation, modeling, and evaluation. Data visualization analysis indicates that OVO was the most frequently used e-wallet top-up service (49%), followed by ShopeePay and GoPay, while i-Saku recorded the lowest usage. Transaction activity peaked at 8:00 PM and was lowest at 4:00 AM. Based on the RFM modeling results, the majority of customers exhibited high recency values (indicating recent transactions) but relatively low monetary values. Using the K-Means clustering algorithm, customers were classified into three clusters: Cluster 0 (Silver): Customers who rarely conduct transactions. Cluster 1 (Gold): Customers who are moderately active and consistent. Cluster 2 (Platinum): Highly active customers with high value for the bank. The evaluation metrics demonstrate that the K-Means algorithm successfully generated accurate, high-quality segmentation, providing a clear representation of customer behavior during e-wallet top-up transactions. The resulting customer segmentation can serve as a strategic reference for determining targeted marketing approaches based on cluster characteristics. Additionally, the findings may assist the bank in identifying optimal timing for promoting other banking services.

REFERENCES

- [1] J. Heikal, "Customer Segmentation in E-Commerce Using the RFM Model : A Case Study of E-Commerce Transaction Data (January 2009 – December 2011) From Kaggle," vol. 9, no. 2017, pp. 7790–7796, 2025.
- [2] M. F. Fadhillah, A. Lovely, A. Suyoso, and I. Puspitasari, "Customer Segmentation with Clustering Algorithm Based on Recency , Frequency , and Monetary (RFM) Attributes Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency , Frequency dan Monetary (RFM)," vol. 5, no. January, pp. 48–56, 2025.
- [3] K. Tsipitsis and A. Chorianopoulos, *Data Mining Techniques in CRM: Inside Customer Segmentation*.
- [4] D. Prasetyo, F. Mahardika, Hariyadi, Nurhayati, Kurniabudi, A. Permadi, J. D. Pratama, Munandar, N. T. Hadi, L. Suhery, N. Maulana, A. I. Lubis, Ilham, A. Kristiyanto, dan M. A. Jabbar, *Cloud Computing*. Yogyakarta, Indonesia: PT Penamuda Media, 2024.
- [5] Y. Syahra *et al.*, "Customer Segmentation Using RFM and K- Means Clustering to Support CRM in Retail Industry," vol. 9, no. 3, pp. 1120–1131, 2025.
- [6] F. Ranja, E. B. Nababan and A. Candra, "Synthetic Data Generation Using Time-Generative Adversarial Network (Time-GAN) to Predict Cash ATM," *2023 International Conference on Computer, Control, Informatics and its*

- Applications (IC3INA)*, Bandung, Indonesia, 2023, pp. 418-423, doi: 10.1109/IC3INA60834.2023.10285809
- [7] M. Alenezi, "An Ontological Framework for Healthcare Web Applications Security," vol. 12, no. 6, pp. 511–516, 2021.
- [8] D. A. Awaliyah, B. Prasetyo, R. Muzayanah, and A. D. Lestari, "Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm," vol. 11, no. 2, pp. 539–548, 2024, doi: 10.15294/sji.v11i2.6137.
- [9] Y. K. Febrina, R. Saputra, K. F. G. M. Informatika, T. Informasi, and P. N. Padang, "E ISSN : 2809-4069 Segmentasi Pelanggan Toko Hanifah Berdasarkan Analisis RFM dengan Metode K-Means Clustering," vol. 5, no. 2, pp. 274–282, 2025.