International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 5, No. 1, March 2026, pp. 1∼8
ISSN: 2828-5611
DOI: 10.30812/IJECSA.v5i1.6087

Application of RFM Model and K-Means Algorithm in Customer Loyalty Segmentation of Indonesian Regional Water Utility Company

Hairani Hairani, Rizki Wahyudi, Gede Yogi Pratama, M. Khaerul Ihsan
Universitas Bumigora, Mataram, Indonesia

Article Info

Article history:

Received January 7, 2026
Revised January 29, 2026
Accepted February 20, 2026

Keywords:

Customer Loyalty
K-Means
PDAM
RFM

ABSTRACT

Customer loyalty is an important factor in sustaining the Regional Drinking Water Company (PDAM) business, as it directly affects customer retention and company revenue. However, PDAM faces challenges in understanding the diverse patterns of customer loyalty. Therefore, this study aims to classify customer loyalty using Recency, Frequency, and Monetary (RFM) analysis and the K-means clustering method. The methods employed are RFM and K-Means, applied to 20,000 PDAM customer transaction records. The results of the study show that K-Means applied to RFM data produces three customer loyalty clusters: high-value customers (Cluster 0) with 4900 instances, medium-value customers (Cluster 2) with 10,600 instances, and low-value customers (Cluster 1) with 4500 instances. The results of the analysis show that the high-value cluster comprises customers who have recently made transactions, have high purchase frequency, and spend high purchase amounts. On the other hand, the low-value cluster comprises customers with low recency, frequency, and monetary values, indicating low loyalty. This study implies that PDAM customer segmentation using RFM analysis and the K-Means algorithm produces three loyalty clusters, namely high value, medium value, and low value, which allow the company to design more targeted service and marketing strategies according to the characteristics of each customer group.

Corresponding Author:

Hairani Hairani,
Computer Science Study Program, Faculty of Engineering, Universitas Bumigora,
Email: hairani@universitasbumigora.ac.id

How to Cite: H. Hairani, R. Wahyudi, G. Y. Pratama, and M. K. Ihsan, Application of RFM Model and K-Means Algorithm in Customer Loyalty Segmentation of Indonesian Regional Water Utility Company, *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 5, no. 1, pp. 1–8, Mar. 2026. doi: 10.30812/ijecsa.v5i1.6087.

1. INTRODUCTION

Water is a basic human need, the fulfillment of which has been regulated in Law No. 7 of 2004 concerning Water Resources, in which regional governments have an obligation to provide clean water services through Regionally-Owned Enterprises (BUMD), one of which is the Regional Drinking Water Company (PDAM) [1]. As a public service entity, PDAM has a significant responsibility not only in providing clean water but also in maintaining customer satisfaction and loyalty [2]. PDAM has a large amount of customer transaction data, but the use of this data has not been optimal to understand the level of customer loyalty comprehensively [3]. This condition hinders the company from formulating targeted service and marketing strategies [4, 5]. In fact, the level of customer loyalty is a crucial indicator in maintaining the continuity of service and financial stability of the company, especially in the era of digital transformation, which is marked by the increasing volume and complexity of customer data [6]. Data-based segmentation is an effective approach to understanding customer behavior and characteristics [7]. One of the most widely used techniques is Recency, Frequency, and Monetary (RFM) analysis, which evaluates the last time a customer made a transaction (Recency), the frequency of transactions (Frequency), and the total value of transactions (Monetary) [8]. RFM is considered capable of providing a comprehensive picture of the value and customer involvement in the services provided [9]. To improve segmentation accuracy, the RFM method can be combined with clustering algorithms such as K-Means [10]. Clustering is a data mining technique that aims to group objects based on similar characteristics [11]. K-Means is known for its efficiency [12].

Segmentation or grouping of customers in a company can help the company in making decisions regarding certain customer groups [13]. Previous studies have demonstrated the effectiveness of the RFM and clustering approaches in various business contexts, including the SME, e-commerce, and public service sectors [14]. Research [15] conducting customer loyalty using the K-Means method for clustering clean water company customers based on water usage volume. Previous studies have not applied the RFM approach; consequently, they have not been able to comprehensively measure customer loyalty, and the clustering results obtained have not been used to design marketing strategies or customer services aligned with PDAM needs. Research [16] applied customer segmentation using the RFM model and the K-Means algorithm, resulting in four clusters: Platinum (38.4% or 557 customers), Gold (12.6% or 184 customers), Silver (22.7% or 330 customers), and Bronze (26% or 378 customers). However, this study has not considered the late payment attribute, which can provide a more accurate picture of customer loyalty.

Research [17] applied the RFM model and K-Means algorithm for PDAM customer segmentation, but have not integrated the public service context, used an analysis framework such as Cross Industry Standard Process for Data Mining (CRISP-DM), or categorized the segmentation results into loyalty levels such as low, medium, and high, which are relevant for policy making in the subscription-based service sector such as PDAM. Research [18] using the RFM and K-Means methods for customer loyalty segmentation which produces 2 categories, namely potential and non-potential.

Previous studies have applied the RFM model and the K-Means algorithm for PDAM customer segmentation. However, these studies have not considered specific public service attributes, such as late payment, which are important in the context of PDAM. Therefore, this study proposes an RFM model with the K-Means algorithm for PDAM customer segmentation and evaluates cluster quality using the Elbow method, thereby distinguishing it from previous studies. Thus, the purpose of this study is to segment PDAM customers using the RFM model and the K-Means algorithm and to classify customers into loyalty levels: low, medium, and high. Thus, this study contributes to PDAM customer segmentation by combining the RFM model with the K-Means algorithm and including late payment behavior as a public service attribute. This method simplifies the classification of customers into low-, medium-, and high-loyalty levels.

2. RESEARCH METHOD

This research uses the CRISP-DM approach, which can be seen in Figure 1, which consists of six stages, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment [19]. Customer transaction data of 20,000 is used as the basis for analysis. The clustering process is performed after transforming the data into RFM values. The Elbow method is used to determine the optimal number of clusters, after which the K-Means algorithm is applied.

Customer segmentation, which divides customers into groups with similar characteristics, allows companies to tailor their approach to each segment more precisely [20]. The data obtained for this customer segmentation study comprises several variables relevant to the analysis. The first variable is Recency, which measures the time since the customer's last transaction. The second variable is Frequency, which denotes how often the customer transacts within a given time period. The third variable is Monetary, which shows the total purchases or transaction value made by the customer [21]. Data preparation is a critical stage in the CRISP-DM methodology, transforming raw data into a format ready for analysis and modeling, as shown in Figure 2.

Customer analysis models are used to evaluate and group customers based on their transaction behavior. Recency measures the time elapsed since the customer's last transaction, indicating how recently the customer transacted. Frequency to calculate how often

a customer makes a transaction during a certain period. Monetary calculates the total value of money spent by a customer during a certain period, and the value of the customer's financial contribution to the company [22]. This study aims to examine customer behavior patterns using RFM analysis and then apply two clustering algorithms, namely K-Means, to identify customer groups with similar characteristics.

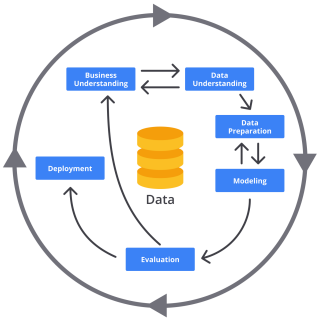

Figure 1. Research Flow [23]

Table 1. Sample Dataset

Customer ID	Water Use	Unit Water Price	...	Late Payment
2313	1	Rp. 16.000	...	8
...	...	...	...	...
2701	611	RP. 13.271.150	...	19

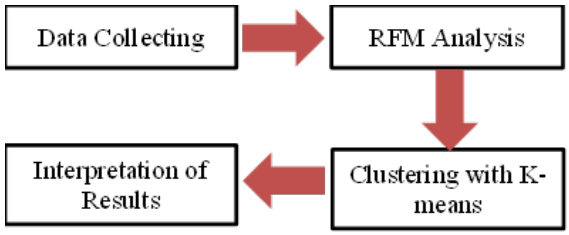

Figure 2. Data Preparation

3. RESULT AND ANALYSIS

In this study, customer segmentation was conducted using the K-means clustering algorithm applied to normalized Recency, Frequency, and Monetary data. The dataset comprised 51,287 rows of customer data, with transaction information randomly sampled from approximately 20,000 data points, which had been calculated and normalized as in Table 2.

Table 2. RFM Analysis Results

No	Customer ID	Recency	Frequency	Monetary
1	32481	3	1	3
2	32501	3	1	3
...	...	...	...	...
19615	110468	2	1	2
19616	110667	2	1	2

In Table 3, the Recency value reflects the length of the customer's delay in making the last transaction, where a value of 1

indicates a customer who has only made a transaction in the last 129 days, a value of 2 for those who have not made a transaction in 3158 days, and a value of 3 for those who have not made a transaction for more than 59 days. Meanwhile, Table 4 describes how often a customer has made a transaction in the last two months: if they made a transaction in only one month, the value is 1; if in both months, the value is 2.

Table 3. Recency Results

No	Recency	Code
1	1-29	3
2	30-59	2
3	>60	1

Table 4. Frequency Results

No	Transaction	Code
1	1 transaction in 1 month	2
2	More than 1 transaction in 1 month	1

Table 5 presents the total amount of payments made by customers in the last two months. Figure 3 shows the results of the analysis using the elbow method to determine the optimal number of clusters. In this graph, the number of clusters is tested from 1 to 10. Each point on the graph represents the distortion (or inertia) value, defined as the sum of squared distances between the data in a cluster and its cluster center. This value is used to assess the quality of the data grouping; the smaller the value, the denser the data within a cluster. The distortion value decreases sharply as the number of clusters increases from 1 to 3. However, thereafter, the decrease is no longer statistically significant. This pattern forms an angle that resembles an elbow, known as the elbow point. This point is considered an indicator of the optimal number of clusters because adding further clusters yields only slight improvements. Based on the graph, the elbow occurs at k = 3, which is used as the optimal number of clusters in this analysis.

Table 5. Monetary Results

No	Recency	Code
1	0-1.000.000	1
2	1.000.000 30.000.000	2
3	>30.000.000	3

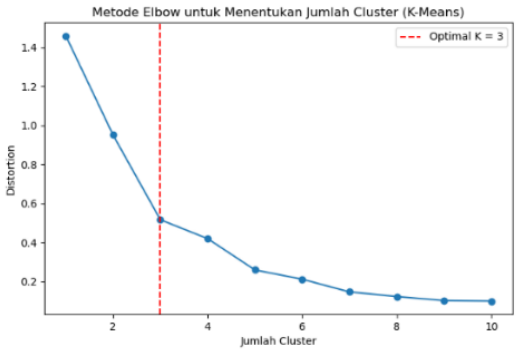

Figure 3. Elbow Method

PDAM customer segmentation is conducted using the K-Means algorithm based on RFM variables to group customers into clusters with similar behavioral characteristics. The clustering results are visualized in Figure 4, which shows the distribution of the number of customers in each cluster. Based on the segmentation results, three main clusters were formed. Cluster 2 has the largest number of customers, exceeding 8,000. Meanwhile, Cluster 0 and Cluster 1 each include approximately 5.800 and 5.500

customers, respectively. The dominance of Cluster 2 shows that the majority of PDAM customers have similar behavioral patterns, which most likely reflect customers with a moderate level of loyalty or ”medium value customer”, when associated with the previous RFM segmentation results.

This distribution provides an important picture for PDAM management in formulating customer service and retention strategies. The group with the largest number (Cluster 2) should be the primary focus of efforts to increase loyalty, for example, through routine education, payment reminders, or loyalty incentives. On the other hand, clusters with a smaller number of customers also need to be analyzed further to find out whether they are very loyal customers (high value) or customers who are at risk of being inactive (low value), so that the approach given can be adjusted to the characteristics of each segment.

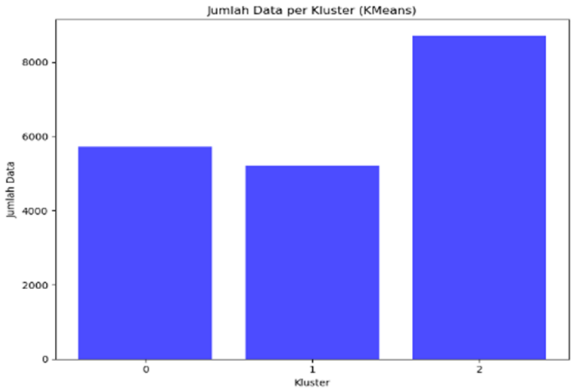

Figure 4. K-Means Method Result

In Figure 5, the segmentation results indicate that most PDAM customers (55.7%) are classified as medium-value customers. This group comprises customers with moderate loyalty; they pay bills regularly but are not yet consistently punctual or consistent in their payment amounts. Customers in this category can be upgraded to more loyal customers through appropriate approaches, such as payment reminders, loyalty programs, or easier access to services. As many as 22.3% of customers are classified as high-value customers, namely a group that is very loyal to PDAM services. They exhibit good payment behavior across three RFM dimensions: paying regularly, paying on time, and paying in relatively large amounts. This group is an important asset to the company and should be retained through retention strategies, such as incentives, special appreciation, or exclusive services, to strengthen customer relationships with PDAM.

Meanwhile, 22.0% of customers are classified as Low Value Customers, indicating the lowest levels of engagement and loyalty. They tend to pay infrequently, delay payments, or only pay in small amounts. This group requires special attention from PDAM, particularly in education and in raising awareness of the importance of regular payments. Strategies such as automated reminders, streamlined payment processes, and incentive programs for timely payments can be used to encourage improved behavior in this group.

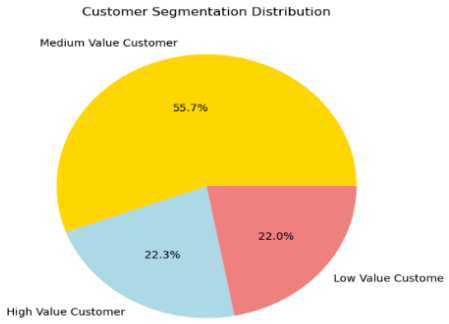

Figure 5. Customer Analysis

4. CONCLUSION

The results of PDAM customer segmentation using the K-Means algorithm and the RFM approach successfully grouped customers into three main clusters: high-value, medium-value, and low-value customers. The high-value cluster comprises customers with active payment behavior and high financial contributions, whereas the medium-value cluster exhibits more stable behavior but less intensive activity than the first cluster. The low-value cluster comprises customers who rarely pay or pay late and make low financial contributions. For future development, it is recommended to compare the segmentation results with those from other clustering methods, such as DBSCAN or Hierarchical Clustering, and to incorporate additional variables, such as customer type or subscription period.

REFERENCES

[1] D. Hadiwijaya and E. Mintarsih, "Pengaruh Penilaian Kerja dan Motivasi Kerja terhadap Kinerja Karyawan pada Perusahaan Daerah Air Minum Tirta Kerta Raharja Kabupaten Tangerang," *JMB : Jurnal Manajemen dan Bisnis*, vol. 10, no. 2, Sep. 28, 2021. DOI: 10.31000/jmb.v10i2.5133.

[2] F. Fetni, S. Baso, and A. R. Septiana, "Analisis Kualitas Layanan Perusahaan Daerah Air Minum pada Pelayanan Air Bersih di Kecamatan Tirawuta Kabupaten Kolaka Timur," *Journal Publicuho*, vol. 6, no. 4, pp. 1496–1510, Dec. 21, 2023. DOI: 10.35817/publicuho.v6i4.293.

[3] M. Effendi, A. D. Mardani, and G. Wijayanto, "The Impact of Digital Service Quality and Consumer Trust on Customer Loyalty with Customer Satisfaction as a Mediator," en, *Management Studies and Entrepreneurship Journal (MSEJ)*, vol. 4, no. 6, pp. 9866–9874, 2023. DOI: 10.37385/msej.v4i6.3913.

[4] M. Z. Zainuddin, S. W. Mahrani, and G. Suprianto, "The Impact of Marketing Management on Customer Loyalty and Company Financial Performance: A Case Study on Retail Companies in Indonesia," en, *Jurnal Ekonomi*, vol. 13, no. 03, pp. 275–280, Jul. 18, 2024.

[5] A. H. Tahir, M. Adnan, and Z. Saeed, "The impact of brand image on customer satisfaction and brand loyalty: A systematic literature review," en, *Heliyon*, vol. 10, no. 16, e36254, Aug. 2024. DOI: 10.1016/j.heliyon.2024.e36254.

[6] M. D. S. Pereira *et al.*, "Factors of Customer Loyalty and Retention in the Digital Environment," en, *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 20, no. 2, p. 71, Apr. 12, 2025. DOI: 10.3390/jtaer20020071.

[7] A. Nugraha *et al.*, "K-Means clustering interpretation using recency, frequency, and monetary factor for retail customers segmentation," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 23, no. 2, p. 435, Apr. 1, 2025. DOI: 10.12928/telkomnika.v23i2.26044.

[8] R. Cristover, H. Toba, and B. R. Suteja, "Segmentation and Formation of Customer Regression Model Based on Recency, Frequency and Monetary Analysis," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, Aug. 26, 2022. DOI: 10.28932/jutisi.v8i2.5075.

[9] F. Safari, N. Safari, and G. A. Montazer, "Customer lifetime value determination based on RFM model," en, *Marketing Intelligence & Planning*, vol. 34, no. 4, pp. 446–461, Jun. 6, 2016. DOI: 10.1108/MIP-03-2015-0060.

[10] H. Hairani *et al.*, "Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 275–282, Mar. 31, 2022. DOI: 10.30812/matrik.v21i2.1542.

[11] A. Iskandar, R. A. A. H. Harahap, and A. G. Ramadhan, "Data Mining Customer Clustering Using K-Means Method," en, *International Journal of Economic Literature*, vol. 2, no. 3, pp. 927–944, Mar. 11, 2024.

[12] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *Jurnal Terapan Teknologi Informasi*, vol. 2, no. 1, pp. 23–32, Apr. 26, 2018. DOI: 10.21460/jutei.2018.21.76.

[13] H. Yusa, M. Sudarma, and N. Pramaita, "HR Potensi Pelanggan Tunggakan PDAM Menggunakan Metode K-Medoids dengan Optimasi Ant Colony Optimization (ACO)," *Majalah Ilmiah Teknologi Elektro*, vol. 17, no. 3, p. 353, Dec. 5, 2018. DOI: 10.24843/MITE.2018.v17i03.P08.

[14] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 2, pp. 229–240, Nov. 11, 2021. DOI: 10.29100/jipi.v6i2.2008.

[15] A. R. Siregar, S. R. Andani, and W. Saputra, "Clustering Pada Pelanggan Perusahaan Air Bersih dengan Algoritma K-Means," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 7–14, Dec. 30, 2020. DOI: 10.30645/brahmana.v1i1.2.

[16] A. P. Pramudiansyah, "Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, pp. 06–19, Sep. 3, 2021. DOI: 10.35329/jiik.v7i2.201.

[17] I. Maskanah, "Segmentasi Pelanggan Toko Purnama dengan Algoritma K-Means dan Model RFM untuk Perancangan Strategi Pemasaran," *INOVTEK Polbeng - Seri Informatika*, vol. 5, no. 2, p. 218, Nov. 27, 2020. DOI: 10.35314/isi.v5i2.1443.

[18] M. H. Anshary, O. Soesanto, and A. Ayatullah, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM (Recency, Frequency, Monetary)," en, *RAGAM: Journal of Statistics & Its Application*, vol. 1, no. 1, pp. 63–73, Dec. 23, 2022. DOI: 10.20527/ragam.v1i1.7382.

[19] N. P. D. T. Yanti and I. M. D. P. Asana, "Sistem Klasifikasi Pengajuan Kredit dengan Metode Support Vector Machine (SVM)," en, *Jurnal Sistem Cerdas*, vol. 6, no. 2, pp. 123–133, Aug. 7, 2023.

[20] Y. A. Singgalen, "Penerapan Metode CRISP-DM untuk Optimalisasi Strategi Pemasaran STP (Segmenting, Targeting, Positioning) Layanan Akomodasi Hotel, Homestay, dan Resort," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, p. 1980, Oct. 25, 2023. DOI: 10.30865/mib.v7i4.6896.

[21] R. Hesananda and P. Apriliga, "Customer Segmentation of Cash Management System Using K-Means Clustering," *Journal of Applied Research In Computer Science and Information Systems*, vol. 2, no. 2, pp. 191–202, Dec. 30, 2024. DOI: 10.61098/jarcis.v2i2.188.

[22] H. Rizkyanto and F. L. Gaol, "Customer Segmentation of Personal Credit using Recency, Frequency, Monetary (RFM) and K-means on Financial Industry," en, *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023. DOI: 10.14569/IJACSA.2023.0140417.

[23] C. Schrfffdfffder, F. Kruse, and J. M. Gfffdfffdmez, "A Systematic Literature Review on Applying CRISP-DM Process Model," en, *Procedia Computer Science*, vol. 181, pp. 526–534, 2021. DOI: 10.1016/j.procs.2021.01.199.

[This page intentionally left blank.]