Jurnal Krisnadana (Jurnal Komputer, Sistem Kendali & Jaringan)
Vol 5, Issue 1, September 2025, pp. 45-56
https://doi.org/10.58982/krisnadana.v5i1.929

E-ISSN: 2808-3563
P-ISSN: 2808-3520

Customer Segmentation Using RFM Model and Fuzzy C-Means at PT SNS 21 Bali

I Made Dwi Putra Asana[1*], Made Leo Radhitya[2], Dewa Nyoman Yogantara[3], I Gede Sudiantara[4], I Putu Noven Hartawan[5]

[1*,2,3]Program Studi Informatika, Institut Bisnis Dan Teknologi Indonesia, Denpasar, Bali, Indonesia
[4,5] Program Studi Bisnis Digital, Institut Bisnis Dan Teknologi Indonesia, Denpasar, Bali, Indonesia

[1]dwiputraasana@instiki.ac.id* ; [2] leo.radhitya@instiki.ac.id ; [3]dewayogantara@gmail.com;
[4]sudiantara@instiki.ac.id; [5]novenhartawan@instiki.ac.id

ARTICLE INFO

Article history:
Received 10 August 2025
Revised 6 September 2025
Accepted 14 September 2025

Keywords:
Customer segmentation;
Recency Frequency Monetary;
Fuzzy C-Means;
CRISP-DM;
Data Mining;

ABSTRACT

This study aims to develop an effective customer segmentation model for PT Sukses Nusantara Sakti 21 Bali, a large-scale multi-level marketing distributor with over 60,000 members. The proposed approach integrates the Recency, Frequency, Monetary (RFM) model with the Fuzzy C-Means (FCM) algorithm to analyze one year of sales transaction data. The CRISP-DM framework was adopted to ensure a structured process, consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Customer transaction records were preprocessed to compute normalized RFM scores, which were then clustered using FCM to capture overlapping membership patterns and better reflect behavioral diversity. The segmentation results were validated using the Silhouette Coefficient and Davies–Bouldin Index, achieving scores of 0.6005 and 0.5093, respectively, indicating high-quality cluster compactness and separation. Three distinct customer segments were identified, each providing actionable insights for targeted marketing strategies, including retention, engagement, and reactivation programs. The findings confirm that integrating RFM and FCM offers a robust and flexible approach for customer segmentation in large-scale MLM contexts. Future work may involve real-time segmentation and integration with predictive analytics to further enhance marketing decision-making.

1. Introduction

Advancements in information technology and digitalization have significantly transformed business strategies, compelling companies to gain deeper insights into customer behavior to enhance retention and business value [1]. PT Sukses Nusantara Sakti 21 Bali (PT SNS 21 Bali), an official distributor of multi-level marketing products in Bali and Nusa Tenggara with over 60,000 members, faces challenges in managing large-scale and heterogeneous customer data. Customer segmentation is a critical strategy for tailoring products, services, and marketing campaigns to customer preferences [2].

The Recency, Frequency, Monetary (RFM) model enables the evaluation of transaction history based on the time of the last purchase, purchase frequency, and purchase value [3]. Meanwhile, the Fuzzy C-Means (FCM) algorithm facilitates the creation of clusters with overlapping memberships, thereby capturing behavioral complexity often overlooked by hard clustering methods such as K-Means [4]. Prior studies have combined RFM and FCM in retail [5], supermarkets [6], and e-commerce [7], yet applications in large-scale distribution with real annual transaction data remain scarce.

This study aims to build and validate a customer segmentation model for PT SNS 21 Bali by integrating the Recency–Frequency–Monetary (RFM) framework with Fuzzy C-Means (FCM). Specifically, we assess whether soft-clustering on normalized RFM can yield business-actionable segments for a large-scale MLM distributor using one year of real transactions, evaluated with the Silhouette Coefficient and Davies–Bouldin Index. Our contributions are threefold: (1) a CRISP-DM–aligned pipeline that transforms raw transaction logs into deployable soft segments; (2) an empirical study on a large member base that demonstrates high-quality clustering (SC = 0.6005; DBI = 0.5093), evidencing compactness and separation; and (3) a practical mapping from the discovered segments to personalized retention, engagement, and reactivation strategies to improve marketing effectiveness.

This study aims to build and validate a customer segmentation model for PT SNS 21 Bali by integrating the Recency–Frequency–Monetary (RFM) framework with Fuzzy C-Means (FCM). Specifically, we assess whether soft-clustering on normalized RFM can yield business-actionable segments for a large-scale MLM distributor using one year of real transactions, evaluated with the Silhouette Coefficient and Davies–Bouldin Index [8], [9]. Our contributions are threefold: (1) a CRISP-DM–aligned pipeline that transforms raw transaction logs into deployable soft segments; (2) an empirical study on a large member base that demonstrates high-quality clustering (SC = 0.6005; DBI = 0.5093), evidencing compactness and separation; and (3) a practical mapping from the discovered segments to personalized retention, engagement, and reactivation strategies to improve marketing effectiveness.

2. Method

A. Research Design

This study employs a data mining approach using the CRISP-DM (Cross Industry Standard Process for Data Mining) model, which consists of six main stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [1]. This method was selected for its ability to provide a structured framework for processing customer transaction data into valuable segmentation insights.

The research framework is illustrated in Fig. 1, which depicts the workflow from data collection to the application of segmentation results for marketing strategy development.

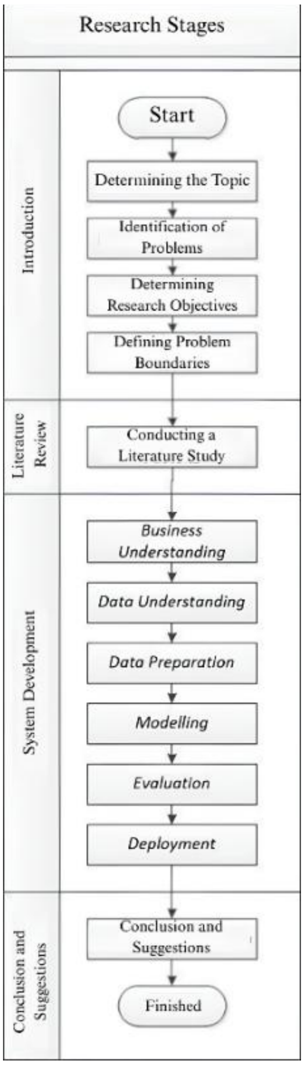

Figure 1. Research Workflow

B. Data collection technique

The dataset for this research consists of:

1. Primary Data:
- Interviews with the owner of Master Stockist PT SNS 21 Bali to gather information on operations, customer profiles, and marketing strategy needs.
- Observation of the sales system to understand the format, structure, and frequency of transaction data.
2. Secondary Data:
- Documentation of sales reports for the period from January 1 to December 31, 2023.
- Literature from books and scientific articles discussing the RFM method and Fuzzy C-Means algorithm [2][3].

A summary of the transaction data used in this study is shown in Table 1.

Table 1. Example of sales transaction data

TRANSACTION_ DATE	CUSTOMER _ID	NO_TRANSACTION	NUMBER OF TRANSACTIONS
Date : 01/01/2023	SNS0043555	10123207152	**702640**
Date : 07/01/2023	SNS0171193	20123207574	**29885440**
Date : 07/01/2023	SNS0203918	20123207567	**71190368**
Date : 07/01/2023	SNS0344243	10123207576	**725088**
Date : 07/01/2023	SNSO322769	20123207585	**5950704**
Date : 07/01/2023	SNS0111813	20123207605	**33954048**
Date : 07/01/2023	SNS0044010	20123207633	**7450880**
Date: 09/01/2023	SNS0314945	10123207701	**412544**
Date: 09/01/2023	SNSOO64391	20123207709	**11901408**
Date: 09/01/2023	SNSO323561	10123207724	**4350528**
Date: 09/01/2023	SNS0323562	10123207729	**4350528**
Date: 09/01/2023	**SNSO322769**	**20123207738**	**5950704**

C. Research Stages

1. Business Understanding

Identify the business needs of PT SNS 21 Bali, particularly in optimizing customer service and retention through accurate segmentation.

2. Data Understanding

Analyze the sales data to assess its patterns, completeness, and quality.

3. Data Preparation

a) Data Transformation: Convert the original PDF sales data into Excel format, then process it using Python with the pandas library.
b) RFM Feature Extraction:

- Recency (R): Number of days since the customer's last transaction relative to a reference date.

Table 2. Recency Result

TRANSACTION_ DATE	CUSTOMER_ ID	NO_TRANSACTION	NUMBER OF TRANSACTIONS	R

Date : 01/01/2023	SNS0043555	10123207152	702640	**8**
Date : 07/01/2023	SNS0171193	20123207574	29885440	**2**
Date : 07/01/2023	SNS0203918	20123207567	71190368	**2**
Date : 07/01/2023	SNS0344243	10123207576	725088	**2**
Date : 07/01/2023	SNSO322769	20123207585	5950704	**2**
Date : 07/01/2023	SNS0111813	20123207605	33954048	**2**
Date : 07/01/2023	SNS0044010	20123207633	7450880	**2**
Date: 09/01/2023	SNS0314945	10123207701	412544	**0**
Date: 09/01/2023	SNSOO64391	20123207709	11901408	**0**
Date: 09/01/2023	SNSO323561	10123207724	4350528	**0**
Date: 09/01/2023	**SNS0323562**	**10123207729**	**4350528**	**0**

- Frequency (F): Number of transactions during the study period.

Table 3. Frequency Result

TRANSACTION_ DATE	CUSTOMER_ ID	NO_TRANSACTION	NUMBER OF TRANSACTIONS	R	F
Date : 01/01/2023	SNS0043555	10123207152	702640	8	**1**
Date : 07/01/2023	SNS0171193	20123207574	29885440	2	**1**
Date : 07/01/2023	SNS0203918	20123207567	71190368	2	**1**
Date : 07/01/2023	SNS0344243	10123207576	725088	2	**1**
Date : 07/01/2023	SNSO322769	20123207585	5950704	2	**2**
Date : 07/01/2023	SNS0111813	20123207605	33954048	2	**1**
Date : 07/01/2023	SNS0044010	20123207633	7450880	2	**1**
Date: 09/01/2023	SNS0314945	10123207701	412544	0	**1**
Date: 09/01/2023	SNSOO64391	20123207709	11901408	0	**1**
Date: 09/01/2023	SNSO323561	10123207724	4350528	0	**1**
Date: 09/01/2023	**SNS0323562**	**10123207729**	**4350528**	**0**	**1**

- Monetary (M): Total purchase value during the study period.

Table 4. Monetary Result

TRANSACTION_ DATE	CUSTOMER_ ID	NO_TRANSACTION	NUMBER OF TRANSACTIONS	R	F	M
Date : 01/01/2023	SNS0043555	10123207152	702640	8	1	**702640**
Date : 07/01/2023	SNS0171193	20123207574	29885440	2	1	**29885440**
Date : 07/01/2023	SNS0203918	20123207567	71190368	2	1	**71190368**
Date : 07/01/2023	SNS0344243	10123207576	725088	2	1	**725088**
Date : 07/01/2023	SNSO322769	20123207585	5950704	2	2	**11901408**
Date : 07/01/2023	SNS0111813	20123207605	33954048	2	1	**33954048**
Date : 07/01/2023	SNS0044010	20123207633	7450880	2	1	**7450880**
Date: 09/01/2023	SNS0314945	10123207701	412544	0	1	**412544**
Date: 09/01/2023	SNSOO64391	20123207709	11901408	0	1	**11901408**
Date: 09/01/2023	SNSO323561	10123207724	4350528	0	1	**4350528**
Date: 09/01/2023	**SNS0323562**	**10123207729**	**4350528**	**0**	**1**	4350528

c) Data Normalization: Normalize RFM values to a uniform scale before applying the FCM algorithm.

Table 5. Normalization Result

R	F	M
1	0	**0.004099**
0,25	0	**0.416414**
0,25	0	**1**
0,25	0	**0.004416**
0,25	1	**0.162323**
0,25	0	**0.473898**
0,25	0	**0.099443**
0	0	**0**
0	0	**0.162323**
0	0	**0.055639**
0	0	**0.055639**
0	**0**	**0.078247**

4. Modelling

The modelling process uses the Fuzzy C-Means (FCM) algorithm, which allows each customer to belong to multiple clusters with varying membership degrees.

The FCM algorithm steps include:

- Setting the number of clusters (c) and fuzziness parameter (m).
- Initializing the membership matrix randomly.
- Calculating cluster centers based on weighted memberships.
- Updating the membership matrix until convergence or maximum iterations are reached [4].

$$u = \frac{1}{\left(\left(\frac{|[1,0,0.004099]-c_1|}{|[1,0,0.004099]-c|}\right)^{\frac{2}{m-1}} + \left(\frac{|[1,0,0.004099]-c_2|}{|[1,0,0.004099]-c|}\right)^{\frac{2}{m-1}}\right)} \quad (1)$$

$$u = \frac{1}{\left(\left(\frac{|[1,0,0.004099]-[0.5,0.5,0.002]|}{|[1,0,0.004099]-[0.6,0.6,0.0015]|}\right)^{\frac{2}{m-1}} + \left(\frac{|[1,0,0.004099]-[0.8,0.8,0.001]\,|}{|[1,0,0.004099]-[0.6,0.6,0.0015]|}\right)^{\frac{2}{m-1}}\right)} \quad (2)$$

Description of Equations:

- Equation (1) represents the Fuzzy C-Means (FCM) formula for calculating the membership degree (u). One of the data points from the previously normalized table is used as a sample calculation. In this example, it is assumed that there are two cluster centers (c_1 and c_2) and the fuzziness parameter (m) is set to 2.
- Equation (2) represents the step following the substitution of the cluster center values (c). In this example, it is assumed that c_1= [0.5,0.5,0.002], c_2= [0.8,0.8,0.001], and c = [0.6,0.6,0.0015]. Based on these assumptions, the equation is obtained as follows.

5. Evaluation

Cluster quality was evaluated using two metrics:

- Silhouette Coefficient (SC): Measures cohesion and separation between clusters.
- Davies-Bouldin Index (DBI): Measures the ratio of intra-cluster to inter-cluster distances, where a lower value indicates better performance [5][6].

6. Deployment

The final segmentation results are used to design personalized marketing strategies, such as special offers for potential customers or loyalty programs for high-value customers.

3. Result and Discussions

A. Data Preparation

The one-year sales transaction dataset of PT SNS 21 Bali (January 1 – December 31, 2023) was transformed into Recency, Frequency, and Monetary (RFM) values as described in Section II. The computation results for each variable are presented in Tables 2–4, while the normalized RFM dataset is shown in Table 5 of the Method section. These preprocessing steps ensured data readiness for the clustering process.

B. Clustering Results

The normalized RFM dataset was processed using the Fuzzy C-Means (FCM) algorithm. Based on experimental testing and evaluation metrics, the optimal number of clusters was determined to be three.
The resulting segments are as follows:

- Cluster 1: High-value customers with high purchase frequency and large transaction amounts.
- Cluster 2: Medium-value customers with moderate activity and potential for growth.
- Cluster 3: Low-value customers with minimal activity.

Table 6. Cluster Result

Member ID	*Recency*	*Frequency*	*Monetary*	*Cluster*
SNS0043555	40	16	8507244	**0**
SNS0171193	5	60	987382151	**0**
SNS0203918	3	66	2584572081	**0**
SNS0344243	179	2	1450176	**1**
SNS0322769	2	122	1300389944	**0**
SNS0111813	3	50	1729754541	**0**
SNS0044010	1	56	425275015	**0**
SNS0314945	11	20	11901408	**0**
SNS0064391	3	61	403299259	**0**
SNS0323561	27	16	69608448	**0**
SNS0323562	20	10	43505280	**0**
SNS0044460	2	11	20271888	**0**
SNS0345633	172	4	2900352	**1**
SNS0341850	159	5	19052288	**1**
SNS0075327	5	13	48576704	**0**
SNS0075328	**356**	**1**	**362544**	**2**

C. RFM-Based Cluster Analysis

Each cluster was analyzed according to the individual RFM dimensions:

- Recency: Cluster 1 has the lowest average recency (recent transactions), while Cluster 3 has the highest.

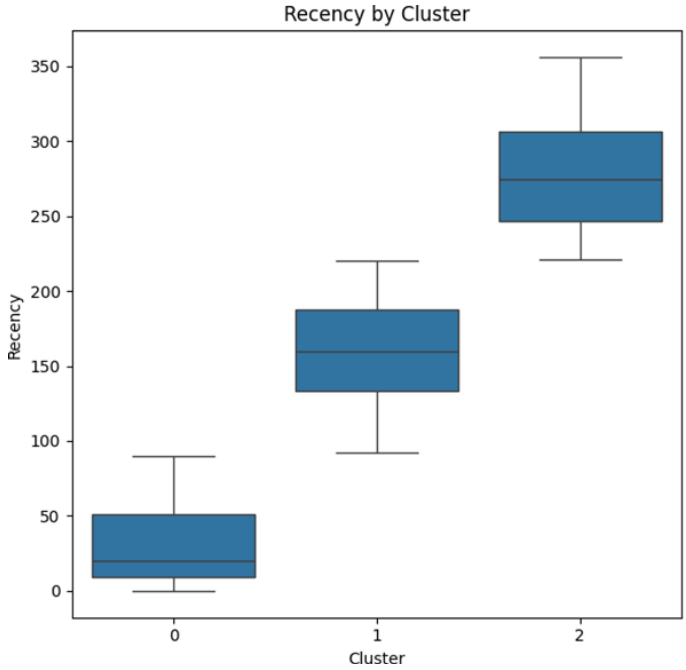

Figure 2. Recency by Cluster

- Frequency: Cluster 1 records the most frequent purchases, Cluster 3 the fewest.

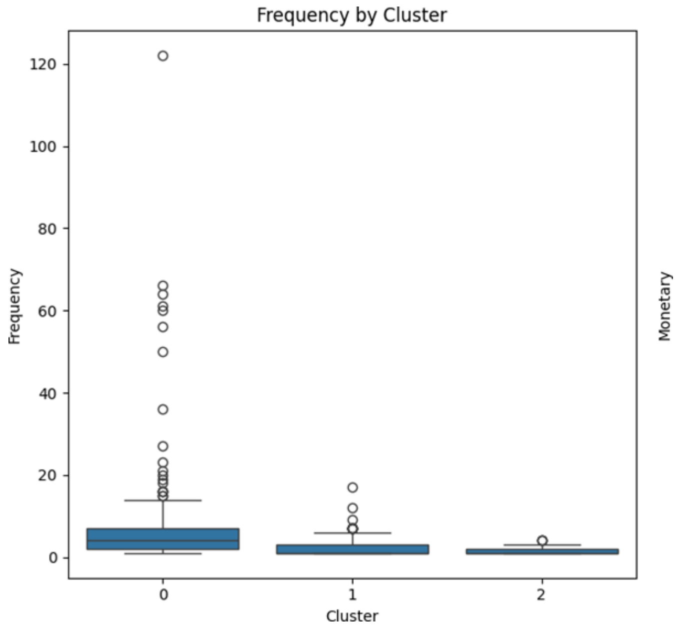

Figure 3. Frequency by Cluster

- Monetary: Cluster 1 contributes the highest total spending, Cluster 3 the lowest.

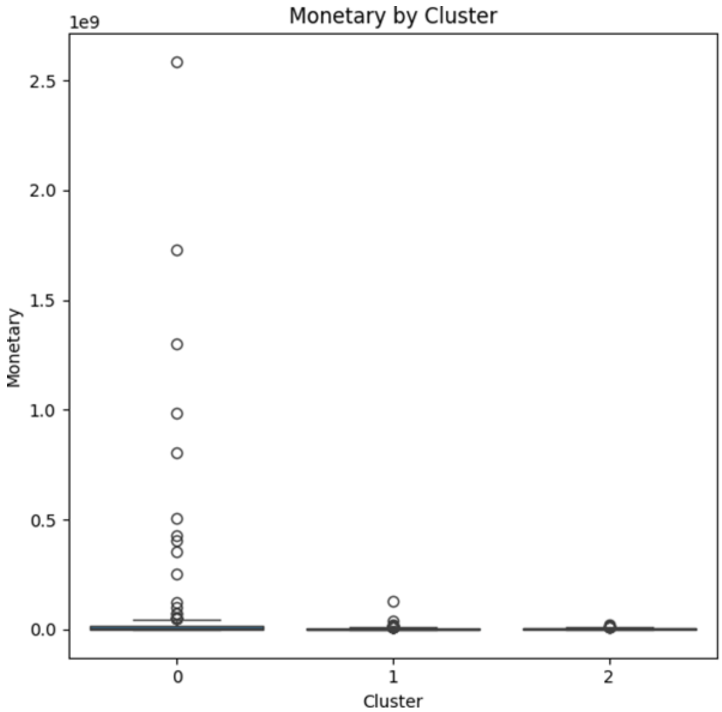

Figure 4. Monetary by Cluster

D. Cluster Validation

The clustering quality was assessed with two internal validity metrics: the Silhouette Coefficient (SC) and the Davies–Bouldin Index (DBI). The SC of 0.6005 indicates good intra-cluster cohesion and inter-cluster separation, while the DBI of 0.5093 reflects compact clusters with minimal overlap. Taken together—a relatively high SC and a low DBI—these results demonstrate that the chosen clustering configuration is effective for this dataset.

E. Discussion

The integration of RFM and Fuzzy C-Means (FCM) produced three business-actionable customer segments with strong internal validity (Silhouette Coefficient = 0.6005; Davies–Bouldin Index = 0.5093), indicating adequate between-cluster separation and within-cluster compactness. Cluster 1 (High Value) displays high recency, frequency, and monetary profiles and should be safeguarded through loyalty programs, exclusive benefits, and personalized offers to preserve lifetime value. Cluster 2 (Potential) shows recent activity with moderate frequency and monetary levels; FCM's soft memberships reveal borderline members partially affiliated with High Value, making them prime candidates for targeted engagement and cross/upsell recommendations. Cluster 3 (Low Value) is characterized by long recency and low purchase intensity, warranting reactivation strategies such as time-bounded discounts, bundles, or win-back campaigns to mitigate churn. Methodologically, FCM is preferable to hard partitioning in this MLM context because it captures overlapping behavior and provides calibrated membership degrees that support intervention prioritization. Practically, the resulting segmentation offers a deployable basis for mapping personalized loyalty, engagement, and reactivation tactics to each segment.

4. Conclusion

This study demonstrated the efficacy of integrating the Recency, Frequency, Monetary (RFM) model with the Fuzzy C-Means (FCM) algorithm for large-scale customer segmentation in the MLM distribution context of PT SNS 21 Bali. The approach yielded three well-defined clusters, validated by a Silhouette Coefficient of 0.6005 and a Davies–Bouldin Index of 0.5093, indicating optimal compactness and separation. The results underscore the model's capability to capture behavioral heterogeneity through soft clustering, enabling nuanced marketing strategies for retention, engagement, and reactivation. This method extends beyond the limitations of hard clustering, offering higher representational fidelity to real-world purchasing patterns. Future research may incorporate multi-dimensional behavioral features and real-time analytics to further enhance segmentation precision and business impact.

Acknowledgment

The authors would like to express their sincere gratitude to the Directorate of Research and Community Service (Direktorat Riset dan Pengabdian Masyarakat) of INSTIKI for the funding support through Decision Letter No. 005/INSTIKI.R1.D2/PM.03/01.2025 and Contract No. 068/INSTIKI.R1.D2/PM.03/01.2025 for the research project titled Customer Segmentation Using the RFM Model and Fuzzy C-Means Algorithm at Master Stokist PT SNS 21 Bali. The authors also thank PT Sukses Nusantara Sakti 21 Bali for providing access to the sales transaction data used in this research.

References

[1] R. Shah Putra, S. A. Sumarni, and R. R. Febriani, "Customer Segmentation Using RFM Model and Fuzzy C-Means in Retail Stores," International Journal of Computer Applications, vol. 176, no. 36, pp. 16–21, Jul. 2020.

[2] N. P. Bulan, "Customer Segmentation Analysis in Supermarkets Using LRFM and K-Means," Journal of Applied Intelligent Systems, vol. 3, no. 2, pp. 75–83, Oct. 2020.

[3] D. Wicaksono, M. I. Wibowo, and N. A. P. Sari, "Implementation of LRFM and Fuzzy C-Means Clustering on E-Commerce Customer Data," Jurnal Teknologi dan Sistem Komputer, vol. 9, no. 1, pp. 15–21, Jan. 2021.

[4] A. Kumar and R. Saha, "Fuzzy Clustering in Big Data Analytics: Techniques and Trends," IEEE Access, vol. 8, pp. 88775–88790, 2020.

[5] H. Han, W. Zhang, and J. Xu, "An Improved RFM Model for Customer Value Evaluation," Procedia Computer Science, vol. 199, pp. 153–160, 2022.

[6] L. Xu, Y. Li, and H. Yu, "Customer Segmentation Based on Improved RFM Model and K-Means Algorithm," IEEE Access, vol. 9, pp. 13721–13728, 2021.
[7] M. I. Alhassan and R. Asare, "Enhancing Customer Segmentation Using Hybrid RFM-Fuzzy Clustering Approach," Journal of Retailing and Consumer Services, vol. 63, p. 102680, May 2021.
[8] Y. Zhao, X. Chen, and Z. Li, "A Comparative Study of Hard and Soft Clustering for Market Segmentation," Expert Systems with Applications, vol. 193, p. 116442, Mar. 2022.
[9] P. Singh and M. K. Sharma, "Application of Soft Computing Techniques in Market Segmentation," Journal of Intelligent & Fuzzy Systems, vol. 41, no. 5, pp. 5377–5388, 2021.
[10] K. T. Chou, "Customer Retention Strategies Based on RFM Segmentation," Asia Pacific Journal of Marketing and Logistics, vol. 34, no. 2, pp. 285–302, 2022.
[11] M. N. Islam, "Soft Clustering for Customer Segmentation: FCM versus GMM," International Journal of Data Science and Analytics, vol. 12, pp. 413–427, 2021.
[12] S. Lee, J. Park, and K. Kim, "RFM-Based Target Marketing Strategies: A Case Study," Journal of Business Research, vol. 138, pp. 123–134, 2022.
[13] X. Wang and Z. Ma, "RFM Analysis for Customer Value Segmentation in E-Business," Procedia Computer Science, vol. 207, pp. 78–85, 2022.
[14] S. Mittal, A. D. Singh, and R. Kaur, "Hybrid RFM and Machine Learning Model for Customer Segmentation," Expert Systems, vol. 39, no. 3, e12865, 2022.
[15] F. Ali, N. Ullah, and M. I. Qureshi, "Enhancing Customer Relationship Management Through Intelligent Segmentation," IEEE Transactions on Engineering Management, vol. 69, no. 6, pp. 3418–3431, Dec. 2022.
[16] Z. Chen and Y. Xu, "RFM-Based Customer Segmentation Using Soft Clustering in Retail Industry," IEEE Access, vol. 10, pp. 25502–25513, 2022.
[17] R. T. Prasetyo and M. R. Hidayat, "Comparative Analysis of K-Means and Fuzzy C-Means in Customer Segmentation," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 6, no. 4, pp. 682–689, Aug. 2022.
[18] S. K. Ghosh, A. Sarkar, and P. Kumar, "A Fuzzy Clustering Approach for Multi-Attribute Customer Segmentation," Applied Soft Computing, vol. 127, p. 109440, May 2022.
[19] A. Sharma, A. K. Sinha, and S. N. Singh, "Advanced RFM Analysis for Customer Segmentation in E-Commerce," Procedia Computer Science, vol. 218, pp. 1804–1813, 2023.
[20] H. Park and Y. Choi, "Personalized Marketing Strategies Using Fuzzy C-Means and RFM Analysis," IEEE Access, vol. 11, pp. 50213–50225, 2023.