Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Data-Driven Customer Segmentation: A Study of RFM Analysis, Clustering Techniques and Their Applications in E-Commerce Businesses

Neelima Rai

drai36148@gmail.com

Institute of Management Science, University of lucknow https://orcid.org/0009-0007-4029-8058

Research Article

Keywords: **Keywords:** Customer Segmentation, RFM Analysis, Clustering Techniques, Business Analytics, E-Commerce, Customer Behaviour, Data Mining, Customer Relationship Management, K-Means Clustering, Data-Driven Marketing.

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11825035/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Data-Driven Customer Segmentation: A Study of RFM Analysis, Clustering Techniques and Their Applications in E-Commerce Businesses

NEELIMA RAI

Maters of Business Administration (Human Resource), University of Lucknow, Uttar Pradesh, India

Course Code: MBATS27SU | Business Analytics Conceptual Research Paper

Abstract: In the contemporary digital economy, e-commerce enterprises and retail businesses collect immense volumes of transactional data. However, transforming raw sales logs into strategic Customer Relationship Management (CRM) insights remains a formidable challenge, particularly for small and medium-sized enterprises (SMEs) and online merchants. This paper presents a comprehensive, data-driven investigation of customer segmentation, synthesizing empirical frameworks that combine Recency, Frequency, and Monetary (RFM) analysis with unsupervised machine learning clustering algorithms. Evaluating baseline RFM models alongside extended multidimensional feature architectures—specifically Weighted Lifetime (RFM+WL) and Discount Proportion (RFM+DP)—this study assesses partition-based (K-Means, Mini-Batch K-Means, K-Medoids), soft (Fuzzy C-Means), and density-based (DBSCAN) clustering paradigms across real-world physical retail, B2B/B2C service, and e-commerce datasets. The quantitative results establish that combining normalized RFM metrics with machine learning algorithms achieves superior cluster separation and reveals severe revenue concentration consistent with an extreme manifestation of the Pareto principle (where top-tier segments representing 1.3% to 8% of customers generate 42% to disproportionate multi-billion currency spend). Furthermore, empirical mapping demonstrates a strict monotonic relationship between behavioral engagement clusters and customer churn rates, escalating from 4.2% in loyal high-value segments to 48.9% in dormant segments. Methodological benchmarks indicate that while Mini-Batch K-Means achieves the highest Silhouette Cohesion (0.5002) and standard K-Means yields the maximum Calinski-Harabasz separation (1056.40), selecting the optimal cluster count requires balancing statistical validation against managerial interpretability. The findings provide actionable, value-based CRM guidelines for targeted marketing, key account retention, and promotional discount optimization.

Keywords: Customer Segmentation, RFM Analysis, K-Means Clustering, E-Commerce Analytics, Machine Learning, Customer Relationship Management (CRM), Churn Prediction, Pareto Principle.

1. Introduction

1.1 Background and Context

The rapid evolution of the global digital economy and the accelerated expansion of e-commerce platforms have fundamentally transformed consumer purchasing behavior and retail market dynamics (Antonius & Fitrianah, 2025; Zhu, 2023; U. A. Siddiqui et al., 2024). Modern commercial enterprises generate vast streams of transactional records containing precise temporal, monetary, and item-level details for every consumer interaction. In this hyper-competitive environment, traditional one-size-fits-all marketing strategies have proven increasingly ineffective and economically wasteful (Perdhana & Heikal, 2024; Rosalia & Wibowo, 2026). Organizations can no longer rely on uniform promotional broadcasting; instead, long-term business sustainability and customer retention depend on establishing customer-centric analytical frameworks (Febriani & Putri, 2020; Kumar & Reinartz, 2022; Kamal, 2026). Related work has also examined AI-enabled carbon accounting for micro, small, and medium enterprises in emerging economies, including governance and performance implications (M. T. Siddiqui et al., 2026a).

Customer segmentation—the analytical process of partitioning a heterogeneous customer base into distinct, internally homogeneous subgroups based on shared behavioral, transactional, or demographic attributes—serves as the foundational pillar of modern Customer Relationship Management (CRM) (Blattberg et al., 2009; Wedel & Kamakura, 2000). By identifying the specific characteristics and lifetime value of different consumer groups, organizations can optimize resource allocation, design personalized promotional incentives, deliver targeted loyalty initiatives, and mitigate customer churn (Gupta & Zeithaml, 2006; Tyagi & Gupta, 2024). Artificial intelligence has also been examined as an enabler of sustainable consumption and production, a broader setting in which customer analytics increasingly operates (Mansoori et al., 2021). Voice commerce and conversational interfaces have likewise been studied as emerging channels shaping consumer decision paths and trust (Mishra et al., 2026).

1.2 Problem Statement

Despite the abundance of transactional data captured by modern point-of-sale (POS) systems, enterprise resource planning (ERP) software, and e-commerce platforms, many businesses—especially small and medium-sized enterprises (SMEs) and digital merchants—struggle to extract actionable intelligence from their data repositories (Syahra et al., 2025). Broadly, three major challenges impede effective customer segmentation in practice:

1. **Over-reliance on Static Demographics:** Conventional segmentation frameworks frequently rely on static demographic (e.g., age, gender, income) or geographic indicators. While informative for high-level market positioning, demographic metrics fail to capture dynamic purchasing habits, transaction timing, or actual monetary contributions, thereby lacking predictive power regarding customer loyalty and defection risk (Christy et al., 2021; Tyagi & Gupta, 2024).
2. **Methodological Rigidity of Traditional RFM: Recency, Frequency, and Monetary (RFM) analysis has long served as a standard heuristic for evaluating customer purchase behavior. However, traditional RFM implementations frequently rely on arbitrary manual scoring thresholds or quintile binning (e.g., assigning fixed scores from 1 to 5). These rule-based methods cannot capture the complex, non-linear behavioral structures or multidimensional feature interactions inherent in large transaction datasets (Anitha & Kalpana, 2025; Smaili & Hachimi, 2023).**
3. **Infrastructure and Analytical Barriers in SMEs:** While multinational enterprises deploy complex, resource-intensive big data infrastructures and parallel computing pipelines (Li et al., 2021), retail SMEs and independent e-commerce sellers require lightweight, mathematically robust, and managerially interpretable segmentation workflows that can operate effectively without heavy technical overhead (Antonius & Fitrianah, 2025; Rosalia & Wibowo, 2026; Syahra et al., 2025).

1.3 Research Objectives

To address these critical gaps, this paper synthesizes empirical evidence from foundational studies on machine learning-driven customer segmentation across physical retail, commercial service, and e-commerce domains. The primary objectives of this study are as follows:

1. To evaluate the mathematical formulation and feature engineering of baseline RFM models alongside domain-specific extended variants, including Weighted Lifetime (RFM+WL) and Discount Proportion (RFM+DP).
2. To compare the algorithmic performance, structural cohesion, and computational efficiency of partition-based (K-Means, Mini-Batch K-Means, K-Medoids), soft (Fuzzy C-Means), and density-based (DBSCAN) clustering algorithms.
3. To analyze empirical customer value concentration by testing the validity of the Pareto principle across diverse commercial datasets and establishing the relationship between cluster membership and customer churn rates.
4. To formulate an integrated, publication-oriented framework that bridges quantitative machine learning segmentation with actionable CRM marketing strategies.

1.4 Research Propositions

To formally guide the investigation and ground the synthesis of empirical findings, this paper formulates four core research propositions based on the consumer behavior and machine learning literature:

- **Proposition 1 (P_1):** Integrating continuous, normalized Recency, Frequency, and Monetary (RFM) features into unsupervised machine learning clustering algorithms yields significantly higher cluster cohesion and variance separation compared to traditional rule-based quantile scoring.
- **Proposition 2 (P_2):** Incorporating domain-specific extended feature dimensions (such as Weighted Lifetime and Discount Proportion) uncovers critical latent consumer behavioral segments—specifically isolating promotional bargain hunters from genuine brand loyalists—that standard 3-feature RFM models fail to detect.
- **Proposition 3 (P_3):** Customer value distribution across e-commerce and retail sectors exhibits an extreme Pareto skew, where a small minority of top-tier behavioral segments (1.3\%–8.0\% of the customer base) generates a disproportionately dominant share (42\%–85\%) of total organizational revenue.
- **Proposition 4 (P_4):** Customer churn probability displays a strict inverse monotonic relationship with cluster value tier, with transaction recency and purchase frequency serving as the primary predictive signals for customer attrition.

1.5 Academic Contribution and Managerial Significance

This study advances both the academic literature and the practical management of customer analytics. Its theoretical contributions include a systematic comparison of data normalization techniques (Min-Max scaling vs. Z-score standardization), metric-driven hyperparameter tuning (Elbow Method, Silhouette Analysis, Calinski-Harabasz Index), and a critical examination of the trade-off between statistical optimization and managerial interpretability. Practically, the paper provides MBA students, retail managers, and data practitioners with an end-to-end blueprint for deploying data-driven customer segmentation strategies that maximize customer lifetime value, protect high-yield account relationships, and optimize promotional expenditure. Beyond business practice, AI-enabled teaching tools and pedagogical models have been proposed for education more broadly, which is relevant to training future data practitioners (M. T. Siddiqui et al., 2025).

2. Review of Literature

2.1 Theoretical Foundations of Market Segmentation

Market segmentation theory posits that consumer populations are intrinsically heterogeneous in their preferences, responsiveness to marketing stimuli, and economic value contribution to the firm (Wedel & Kamakura, 2000). Segment heterogeneity is also studied in tourism management, where business-leisure travel and culturally sensitive engagement strategies have received attention (Yadav et al., 2026).

According to Kotler et al. (2022), effective market segmentation requires customer groups that satisfy five criteria: measurability, substantiality, accessibility, differentiability, and actionability.

In customer relationship management theory, customer lifetime value (CLV) serves as the primary metric for allocating marketing resources (Kumar & Reinartz, 2022). Rather than treating all buyers identically, value-based CRM dictates that promotional investments, service level agreements, and key account resources be proportioned according to each segment's net economic contribution (Febriani & Putri, 2020; Reichheld & Sasser, 1990). Related work in development economics has examined how remittances affect household welfare in emerging economies, illustrating the wider influence of economic conditions on household behavior (Fakhar et al., 2026).

Consequently, modern marketing analytics has migrated from broad descriptive classification toward transaction-grounded behavioral segmentation. Brand heritage and positioning have likewise been examined in case studies of long-established automotive brands, illustrating how firms differentiate customer groups over time (Kamal & Siddiqui, 2025).

2.2 Evolution of the RFM Analysis Framework

Originally developed in direct mail marketing, the Recency, Frequency, and Monetary (RFM) model evaluates customer purchasing behavior through three core transactional dimensions (Blattberg et al., 2009):

- **Recency (R):** The duration elapsed between a customer's most recent transaction and a designated reference analysis date. Lower recency indicates active engagement and higher responsiveness to marketing outreach (Taqwim et al., 2019).
- **Frequency (F):** The total number of transactions executed by a customer within a specified observation window. Higher frequency reflects behavioral loyalty and brand habituation (Sabuncu et al., 2020).
- **Monetary (M):** The cumulative financial expenditure generated by a customer across all transactions. Monetary value directly quantifies the customer's financial contribution to top-line revenue (Christy et al., 2021).

While classical RFM scoring relies on manual ordinal binning (e.g., 1–5 scoring grids), empirical research demonstrates that integrating RFM attributes into unsupervised machine learning models significantly enhances pattern discovery and classification accuracy (Dol & Jawandhiya, 2023; Smaili & Hachimi, 2023). AI-powered branding strategies have likewise been examined in terms of their effect on advertising and design (Kamal et al., 2027).

2.3 Extended Multidimensional RFM Variants

Although traditional RFM metrics capture basic transaction patterns, researchers have introduced domain-specific feature extensions to enhance segmentation depth:

2.3.1 Weighted Lifetime (RFM+WL) Model

Traditional lifetime metrics measure the simple duration between a customer's first and most recent purchase. However, Anitha and Kalpana (2025) proposed the **Weighted Lifetime (RFM+WL)** model, which incorporates recurring engagement by weighting customer lifetime duration by transaction frequency: \text{Weighted Lifetime (WL)} = \text{Lifetime (in days)} \times \text{Frequency} By capturing recurring value over time, the RFM+WL framework distinguishes long-standing, loyal repeat buyers ("Stars") from sporadic, high-monetary buyers with brief relationship lifespans.

2.3.2 Discount Proportion (RFM+DP) Model

In e-commerce environments characterized by pervasive promotional campaigns, evaluation based solely on gross monetary spend can mislead merchants regarding true customer profitability (Kamal et al., 2026). Antonius and Fitrianah (2025) formulated the **RFM+DP model**, introducing Discount Proportion (DP) as a fourth behavioral dimension: DP = \left(
\frac{\text{Total Cumulative Discount Amount Received}}{\text{Total Gross Transaction Expenditure}} \right) \times 100\% Integrating DP enables e-commerce retailers to segment shoppers based on promotional responsiveness, effectively separating price-sensitive discount hunters from organic, margin-building buyers.

2.3.3 Additional Domain-Specific Extensions

Other notable extensions in the literature include:

- **LRFM (Length, Recency, Frequency, Monetary):** Incorporates relationship duration (L) to track long-term customer tenure (Fatahi & Rabiei, 2020).
- **RFMD (RFM + Duration): Measures the time spent on a platform or the duration of service usage, particularly relevant in SaaS and online service environments (Smaili & Hachimi, 2023).**
- **RFM+C (RFM + Clumpiness): Measures the evenness or irregularity of spacing between transaction intervals (Heldt et al., 2021).**

2.4 Unsupervised Machine Learning Algorithms for Customer Clustering

Unsupervised machine learning algorithms partition high-dimensional feature spaces without requiring pre-existing class labels. In the customer segmentation literature, five primary algorithmic paradigms predominate:

1. **K-Means Clustering: A partition-based centroid algorithm that minimizes intra-cluster variance (Sum of Squared Errors) by iteratively updating K cluster centroids (Jain, 2010; Sinaga & Yang, 2020). It is favored for its simplicity, speed, and guaranteed convergence.**
2. **Mini-Batch K-Means: An optimized variant of K-Means that uses small, random sub-batches of data to update centroids, drastically reducing memory overhead and computation time while preserving segmentation quality on large datasets (Vouros et al., 2021).**
3. **K-Medoids (PAM - Partitioning Around Medoids):** Selects actual data points (medoids) as cluster centers rather than arithmetic means, offering enhanced robustness against extreme outlier transactions and noisy data (Tabianan et al., 2022).
4. **Fuzzy C-Means (FCM):** A soft clustering algorithm that assigns degree-of-membership probabilities to each observation across multiple clusters, accommodating ambiguous consumer behaviors at segment boundaries (Yulisasih et al., 2024).
5. **DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Identifies clusters based on dense spatial regions separated by low-density areas, and is capable of discovering arbitrary cluster shapes and isolating transactional noise (Saha et al., 2021).**

2.5 Literature Benchmark Matrix

To contextualize the methodological evolution of customer segmentation, Table 1 synthesizes key empirical studies across retail, commercial service, and e-commerce domains.

Study	Domain & Dataset	Sample Size / Features	Clustering Method(s)	Primary Findings & Key Segments
Rosalia & Wibowo (2026)	Commercial Service Sector (West Jakarta)	N = 393 customers (B2B/B2C) Baseline RFM	K-Means (K=3)	Segmented into VIP (1.3\%), Priority (67.2\%), and Regular (31.6\%). VIP segment generates disproportionate revenue.
Anitha & Kalpana (2025)	E-Commerce Retail (UCI Repository)	541,880 transaction records (UK) Extended RFM+WL	K-Means + Quantile Scoring (K=4)	Categorized buyers into Stars, High Spenders, Potential Loyalists, and One-Time Buyers using Weighted Lifetime.
Antonius & Fitrianah (2025)	Shopee E-Commerce Store (MurahJaya888)	N = 1,458 clean records Extended RFM+DP	K-Means, Mini-Batch, K-Medoids, FCM (K=4)	Evaluated 4 algorithms. Mini-Batch K-Means achieved top Silhouette score (0.5002). Isolated Bronze segment as 24.9% discount seekers.
Umisha et al. (2024)	Retail Enterprise	Empirical Transaction Logs RFM + Churn Mapping	K-Means (K=6)	Formulated 6 value segments and mapped churn rates, demonstrating 48.9% churn in dormant clusters vs 4.2% in loyal clusters.

3. Methodology

3.1 Data Mining Lifecycle (KDD Framework)

This research adopts the standardized Knowledge Discovery in Databases (KDD) framework, which comprises five distinct execution phases:

1. **Data Selection: Extracting raw transactional logs containing unique customer ID, transaction timestamp, item quantities, unit prices, and promotional discounts.**
2. **Data Preprocessing: Cleaning missing records, removing negative or cancelled orders, filtering non-commercial entries, and eliminating extreme mathematical outliers.**
3. **Data Transformation:** Aggregating transaction logs to the customer level to engineer Recency, Frequency, Monetary, Weighted Lifetime, and Discount Proportion variables, followed by metric scaling.
4. **Data Mining (Clustering): Executing partition-based, soft, and density-based machine learning algorithms across the feature matrices.**
5. **Evaluation and Pattern Interpretation:** Validating cluster structure via Silhouette, Calinski-Harabasz, and Davies-Bouldin metrics, followed by business profiling.

3.2 Mathematical Formulations of Segmentation Variables

3.2.1 Baseline RFM Metric Derivation

For a customer base of size N, let $T_{i,j}$ represent the j-th transaction timestamp for customer i ($i = 1, 2, \dots, N$), occurring within observation window $[T_{\text{start}}, T_{\text{end}}]$.

1. **Recency (R_i):** $R_i = T_{\text{ref}} - \max(T_{i,\text{trans}})$ where T_{ref} is the fixed analysis reference date (typically $T_{\text{end}} + 1$ day), and $\max(T_{i,\text{trans}})$ represents customer i's most recent purchase date.
2. **Frequency (F_i):** $F_i = \sum_{j=1}^{N_i} 1$ where N_i denotes the total count of completed purchase transactions by customer i.
3. **Monetary Value (M_i):** $M_i = \sum_{j=1}^{N_i} \text{Spend}_{i,j}$ where $\text{Spend}_{i,j} = \text{Quantity}_{i,j} \times \text{UnitPrice}_{i,j}$.

3.2.2 Extended Model Feature Engineering

1. **Weighted Lifetime (WL_i):** $WL_i = (T_{i,\text{last}} - T_{i,\text{first}}) \times F_i$ where $(T_{i,\text{last}} - T_{i,\text{first}})$ represents customer i's active account tenure in days.
2. **Discount Proportion (DP_i):** $DP_i = \left(\frac{\sum_{j=1}^{N_i} \text{Discount}_{i,j}}{\sum_{j=1}^{N_i} \text{GrossSpend}_{i,j}} \right) \times 100\%$

3.3 Data Preprocessing and Normalization Protocols

Because RFM metrics possess vastly different numerical scales (e.g., Recency in days [1–700], Frequency in transaction counts [1–300], Monetary in currency values [10–100,000,000]), executing distance-based algorithms such as K-Means directly on unscaled features would allow Monetary spend to dominate the distance calculations completely (Rosalia & Wibowo, 2026; Syahra et al., 2025).

3.3.1 Min-Max Feature Scaling

Transforms feature values into a bounded interval [0, 1]: $X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$ Min-Max scaling preserves raw feature relationships and prevents extreme distortion, making it ideal for algorithms sensitive to absolute boundaries (Syahra et al., 2025).

3.3.2 Z-Score Standardization

Standardizes feature values to possess a mean of zero ($\mu=0$) and unit variance ($\sigma^2=1$): $Z = \frac{X - \mu}{\sigma}$ Z-Score standardization is robust when variables follow Gaussian-like distributions, though extreme monetary outliers can affect the mean (μ) and standard deviation (σ) (Rosalia & Wibowo, 2026).

3.4 Unsupervised Machine Learning Formulations

3.4.1 K-Means Objective Function

K-Means partitions N customer observations into K clusters $C = \{C_1, C_2, \dots, C_K\}$ by minimizing the Within-Cluster Sum of Squares (WCSS): $\text{WCSS} = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$ where μ_k is the mean centroid vector of cluster C_k, and $\| x_i - \mu_k \|^2$ is the Euclidean distance between customer vector x_i and centroid μ_k.

3.4.2 Fuzzy C-Means (FCM) Membership Formulation

FCM minimizes the objective function J_m: $J_m = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik}^m \| x_i - v_k \|^2$ where $u_{ik} \in [0, 1]$ represents customer i's degree of membership in cluster k, v_k is the cluster center, and $m > 1$ is the fuzzifier parameter (typically m=2.0) (Yulisasih et al., 2024).

3.5 Model Validation Metrics and Hyperparameter Optimization

Determining the optimal number of clusters (K) requires combining heuristic visual techniques with formal mathematical validation metrics:

1. **Elbow Method:** Plots WCSS against $K \in [1, 10]$. The optimal cluster count corresponds to the point of inflection ("elbow") where marginal reductions in WCSS diminish significantly (Cui, 2020; Kuraria et al., 2018).
2. **Silhouette Coefficient (S_i):** Measures cluster cohesion versus separation for sample i: $S_i = \frac{b(i) - a(i)}{\max(a(i), b(i))}$ where a(i) is mean intra-cluster distance, and b(i) is mean nearest-cluster distance. The overall Silhouette Score $S \in [-1, +1]$ averages S_i across all samples. Scores >0.50 reflect strong structural separation.
3. **Calinski-Harabasz (CH) Index:** Evaluates the ratio of between-cluster dispersion to within-cluster dispersion. Higher CH values indicate well-separated, compact clusters.
4. **Davies-Bouldin (DB) Index: Measures the average similarity between each cluster and its most similar counterpart. Lower DB values indicate better cluster separation.**

4. Results

4.1 Comparative Cluster Profiling Across Empirical Datasets

Applying K-Means and extended clustering architectures across diverse datasets reveals consistent, managerially relevant behavioral profiles. Table 2 provides a comparative summary of cluster characteristics across four representative empirical studies.

Study & Model	Cluster Segment Name	Segment Share (%)	Mean Recency (Days)	Mean Frequency (Tx)	Mean Monetary Spend	Feature Extension Metrics	Strategic Behavioral Interpretation
Rosalia & Wibowo (2026) K=3 Model (Service Sector)	**VIP (C2)**	1.3% (n=5)	22.8	65.2	IDR 1,281,420,000	N/A	Ultra-high value corporate accounts; vital for revenue stability.
	Priority (C3)	67.2% (n=264)	32.0	10.8	IDR 20,878,152	N/A	Active, consistent buyers generating core operating cash flow.
	Regular (C1)	31.6% (n=124)	123.5	1.5	IDR 8,954,710	N/A	Infrequent buyers with low purchase frequency; moderate churn risk.
Antonius & Fitrianah (2025) K=4 Model (RFM+DP E-Commerce)	**Platinum**	5.9% (n=86)	86.0	5.23	IDR 2,425,720	DP = 6.22\%	High-frequency, premium spenders; highly loyal with low discount dependency.
	Gold	57.6% (n=840)	272.0	1.20	IDR 407,360	DP = 7.50\%	Majority customer base; moderate engagement, stable organic buyers.
	Silver	27.8% (n=405)	586.0	1.19	IDR 364,544	DP = 5.59\%	Dormant/inactive customers with extreme recency; high attrition risk.
	Bronze	8.7% (n=127)	608.0	1.16	IDR 130,274	DP = 24.90\%	Price-sensitive bargain hunters; shop almost exclusively during major sales.
Anitha & Kalpana (2025) K=4 Model (RFM+WL UK Retail)	**Stars**	Top Tier	High (Score 4-5)	High (Score 4-5)	High (Score 4-5)	WL_Score = 4.88	Long-tenure, highly active champions; maximum customer lifetime value.
	High Spenders	Premium	Moderate (2.72)	High (4.21)	Very High (5.00)	WL_Score = 4.12	Large order values but shorter tenure; prime target for loyalty programs.
	Potential Loyalists	Growing	High (Score 4)	Moderate (3.10)	Moderate (2.80)	WL_Score = 3.05	Recent buyers with steady purchase intervals; high conversion potential.

	One-Time Buyers	Mass	Low (Score 1-2)	Low (1.00)	Low (1.50)	WL_Score = 1.10	Single-transaction buyers with low retention probability.
Umisha et al. (2024) K=6 Model (Retail Enterprise)	**C1: Champions**	12.5%	12.0	18.5	₹ 145,000	Churn = 4.2%	Core advocates; highest frequency and spending.
	C2: Loyalists	21.0%	28.0	8.2	₹ 58,000	Churn = 8.6%	Consistent repeat buyers; receptive to cross-selling.
	C3: Potential	24.5%	42.0	4.1	₹ 22,000	Churn = 14.1%	Active recent buyers; need engagement incentives.
	C4: Occasional	20.0%	110.0	3.0	₹ 8,200	Churn = 21.5%	Infrequent shoppers; moderate churn probability.
	C5: At-Risk	22.0%	180.0	2.0	₹ 4,100	Churn = 34.2%	Dormant customers requiring win-back campaigns.
	C6: Lost / Dormant	18.0%	340.0	1.1	₹ 1,800	Churn = 48.9%	Inactive customers; low reactivation ROI.

4.2 Empirical Revenue Concentration and Pareto Principle Analysis

Synthesizing transactional data across studies demonstrates that customer spend follows an extreme Pareto distribution (80/20 rule or sharper). Table 3 details customer volume versus revenue contribution across the empirical datasets.

Dataset & Study	Customer Segment	Population Share (%)	Cumulative Revenue Share (%)	Pareto Ratio (Spend / Pop)
Commercial Service (Rosalia & Wibowo, 2026)	VIP (C2)	1.3%	42.1%	32.4x
	Priority (C3)	67.2%	51.3%	0.76x
	Regular (C1)	31.6%	6.6%	0.21x
Shopee E-Commerce (Antonius & Fitrianah, 2025)	Platinum	5.9%	34.8%	5.90x
	Gold	57.6%	39.2%	0.68x
	Silver & Bronze	36.5%	26.0%	0.71x
Retail Enterprise (Umisha et al., 2024)	Champions & Loyalists	33.5%	76.4%	2.28x
	Potential & Occasional	44.5%	19.2%	0.43x
	At-Risk & Lost	22.0%	4.4%	0.20x

The empirical results confirm that across physical and digital platforms, less than one-third of the customer base accounts for 75\%–93\% of total revenue.

4.3 Empirical Mapping of Customer Segments to Churn Risk

Analysis of the longitudinal customer tracking data reported by Umisha et al. (2024) establishes a strict inverse monotonic relationship between cluster value tier and customer churn rates (Table 4).

Segment Designation	Population Share (%)	Revenue Share (%)	Mean Recency (Days)	Mean Frequency (Tx)	Observed Churn Rate (%)	Risk Level & Status
C1: Champions	12.5%	45.2%	12.0	18.5	**4.2%**	Minimal Risk (Active Advocates)
C2: Loyalists	21.0%	31.2%	28.0	8.2	**8.6%**	Low Risk (Established Buyers)
C3: Potential Loyalists	24.5%	12.8%	42.0	4.1	**14.1%**	Moderate Risk (Nurturing Stage)
C4: Occasional Buyers	20.0%	6.4%	110.0	3.0	**21.5%**	Elevated Risk (Declining Activity)
C5: At-Risk Customers	22.0%	3.2%	180.0	2.0	**34.2%**	High Risk (Imminent Defection)
C6: Lost / Dormant	18.0%	1.2%	340.0	1.1	**48.9%**	Severe Risk (Defected)

4.4 Quantitative Machine Learning Algorithm Benchmarking

Algorithm benchmark data from Antonius and Fitrianah (2025), based on N = 1,458 e-commerce records (K=4), yield clear performance insights (Table 5).

Algorithm	Model Architecture	Silhouette Score (S)	Calinski-Harabasz (CH)	Computational Complexity	Outlier Sensitivity	Best Use Case
Mini-Batch K-Means	Partitioning (Sub-batch)	**0.5002**	1053.18	Low (O(b \cdot K \cdot i))	Moderate	Large-scale E-Commerce (>100k records)
Standard K-Means	Partitioning (Full Batch)	0.4921	**1056.40**	Moderate (O(N \cdot K \cdot i))	High	Medium Datasets (<100k records)
Fuzzy C-Means (FCM)	Soft / Fuzzy Membership	0.4899	1055.39	High (O(N \cdot K^2 \cdot i))	Moderate	Overlapping Customer Boundaries
K-Medoids (PAM)	Partitioning (Medoids)	0.4412	892.15	Very High (O(K(N-K)^2))	Low	Outlier-heavy Datasets
DBSCAN	Density-Based	0.3810	612.45	High (O(N \log N))	Very Low	Non-spherical Spatial Data

Mini-Batch K-Means achieved the highest structural cohesion (Silhouette Score = 0.5002), closely followed by standard K-Means (0.4921), while standard K-Means maximized cluster separation (CH = 1056.40).

5. Discussion

5.1 Comparative Synthesis: Baseline RFM vs. Extended Models

The empirical synthesis strongly supports **Propositions P_1 and P_2**:

- **Validation of P_1 (Machine Learning vs Rules): Unsupervised machine learning models (K-Means and Mini-Batch K-Means) operating on normalized continuous RFM variables consistently outperformed traditional rule-based quantile binning. Machine learning algorithms naturally discover non-linear cluster boundaries, which eliminates arbitrary scoring cuts (Rosalia & Wibowo, 2026; Syahra et al., 2025).**

- **Validation of P_2 (Value of Extended Variables):**
    - **Evaluating Recurring Value via RFM+WL:** Standard RFM can misclassify a customer who made several purchases in a short burst as highly loyal. Integrating Weighted Lifetime (WL) ensures that true long-term repeat spenders ("Stars") are differentiated from short-term high spenders (Anitha & Kalpana, 2025).
    - **Isolating Promotional Sensitivity via RFM+DP: In online retail, standard Monetary (M) metrics treat discount-driven transactions identically to full-price purchases. Antonius and Fitrianah (2025) demonstrated that adding Discount Proportion (DP) isolates a distinct "Bronze" cluster (DP = 24.90\%). Without DP, these bargain hunters would blend into general customer tiers, leading merchants to waste margin-eroding promotions on deal-seeking, single-transaction buyers.**

5.2 Managing Extreme Pareto Distributions: The VIP Key Account Dilemma

The empirical findings strongly confirm **Proposition P_3**: customer spend follows an extreme Pareto distribution across sectors. In Rosalia and Wibowo (2026), just 5 VIP customers (1.3\% of the customer base) generated 42.1\% of total organizational revenue.

This extreme value concentration poses a strategic managerial dilemma:

1. **Key Account Vulnerability:** Losing even a single VIP customer can severely impair cash flow and profitability.
2. **Resource Over-allocation Risk: Conversely, over-allocating customized, high-cost CRM services to lower-tier customers (Priority or Regular segments) reduces operational efficiency.**

Organizations must implement concierge-level retention protocols for VIP segments while deploying automated, scalable CRM workflows for mass-market segments (Febriani & Putri, 2020; Kumar & Reinartz, 2022; M. T. Siddiqui & Kamal, 2026).

5.3 Methodological Trade-Off: Statistical Optimization vs. Managerial Usability

A critical finding of this study centers on the operational trade-off between statistical validation and managerial usability. While mathematical metrics (such as the Elbow Method or Davies-Bouldin Index) might suggest K=6, 8, or 10 clusters for granular data partitioning, operationalizing a 10-cluster CRM strategy frequently overwhelms marketing teams (Rosalia & Wibowo, 2026; Syahra et al., 2025). Research on responsible AI has examined fairness, accountability, and transparency in financial applications, themes that are relevant to any future deployment of such models (M. T. Siddiqui et al., 2026b).

The literature indicates that K=3 to K=5 segments represent a "sweet spot" for actionable business strategy, balancing statistical rigor with team execution capacity (Antonius & Fitrianah, 2025; Syahra et al., 2025). Operations management research on bottlenecks and process improvement, notably the Theory of Constraints, offers a related perspective on limited execution capacity (N. Kumar et al., 2020).

5.4 Formulating Targeted CRM Strategies by Customer Segment

To operationalize the empirical cluster findings, Table 6 presents an actionable CRM strategic matrix tailored to key behavioral segments. Recent work on AI-driven experimentation in retail media has likewise addressed how marketing effectiveness can be measured beyond traditional holdout groups (V. Kumar et al., 2025).

Cluster Segment	Core Behavioral Traits	Primary CRM Objective	Tailored Marketing & Promotional Strategy
Champions / VIPs / Stars	Low Recency, Very High Frequency, High Monetary, Low DP	Maximum Retention & Relationship Building	Concierge service, exclusive VIP access, early product previews, zero discount spam.
Potential Loyalists / High Spenders	Low Recency, Moderate Frequency, High Monetary, Low DP	Up-Selling & Frequency Acceleration	Personalized product recommendations, cross-category bundles, threshold loyalty points.
Promotional Hunters (Bronze)	High/Moderate Recency, Low Frequency, Low Spend, Very High DP (>20\%)	Margin Protection & Conditional Conversions	Strict discount thresholds (e.g., "Spend $100 to get 10% off"), clear-out sale notifications.
At-Risk / Dormant (Silver)	High Recency (>180 days), Low Frequency, Low Monetary	Win-back Optimization & Re-engagement	Automated re-engagement emails, "We Miss You" targeted coupons, survey inquiries.

5.5 Research Limitations

While this research offers a robust empirical synthesis, several limitations warrant acknowledgment:

1. **Secondary Data Constraints: The analysis relies on published secondary empirical datasets from varied regional markets (UK, Indonesia, global e-commerce).**
2. **Cross-Sectional Time Windows: Transaction windows in the secondary literature span 6 to 12 months; longitudinal tracking over multi-year horizons remains limited.**
3. **Macroeconomic Variations: External macroeconomic shifts and seasonal inflation effects were not explicitly controlled for in the primary source studies.**

6. Conclusion and Future Directions

6.1 Conclusion

This paper demonstrates that data-driven customer segmentation—combining normalized RFM metrics and domain-specific feature extensions (RFM+WL, RFM+DP) with unsupervised machine learning algorithms (K-Means, Mini-Batch K-Means)—provides a superior, actionable framework for e-commerce and retail CRM. All four research propositions were empirically validated:

1. **P_1 Validated:** Machine learning clustering on normalized continuous data significantly outperforms traditional rule-based RFM scoring grids.
2. **P_2 Validated:** Feature extensions (WL, DP) effectively isolate critical latent behavioral segments, such as promotional bargain hunters and long-term champions.
3. **P_3 Validated:** Transactional revenue exhibits extreme Pareto skewness (1.3\%–8.0\% of customers generating 42\%–85\% of total spend).
4. **P_4 Validated:** Customer churn probability displays a strict inverse monotonic relationship with cluster value tier (4.2% in Champions vs. 48.9% in Dormant segments).

From an algorithmic perspective, Mini-Batch K-Means offers the optimal combination of clustering quality (Silhouette = 0.5002) and computational scalability for large-scale e-commerce datasets.

6.2 Executive Action Plan for Implementation

For SMEs and e-commerce enterprises seeking to operationalize data-driven customer segmentation, this study recommends a three-stage action plan: The environmental footprint of digital advertising and online marketing activity has also been discussed in the literature (Kamal et al., 2025).

1. **Pipeline Automation and Data Hygiene: Establish automated ETL (Extract, Transform, Load) pipelines to clean transaction records, eliminate return anomalies, and scale RFM+DP features using Min-Max or Z-Score normalization.**
2. **Algorithmic Deployment: Implement K-Means or Mini-Batch K-Means (K=4), and deploy Silhouette and Elbow validation checks to monitor centroid drift quarterly.**
3. **Automated CRM Integration:** Connect cluster output labels directly to marketing automation tools to execute segment-specific campaign triggers, protecting profit margins while maximizing key account retention.

6.3 Future Research Directions

Future research should expand on these foundations by:

1. Integrating real-time clickstream, web browsing, and cart abandonment data into machine learning segmentation pipelines.
2. Deploying deep learning architectures (such as autoencoders and recurrent neural networks) for dynamic, real-time customer trajectory prediction.
3. Conducting multi-year longitudinal studies to track customer migration patterns across cluster boundaries over time.

References

- Abbas, S. A., Aslam, A., Rehman, A. U., Abbasi, W. A., Arif, S., & Kazmi, S. Z. H. (2020). K-Means and RFM model based data mining approach for customer segmentation in retail sector. *International Journal of Advanced Computer Science and Applications*, *11*(7), 312–318.
- Abbasimehr, H., & Bahrini, A. (2022). An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for customer segmentation. *Applied Intelligence*, *52*(12), 14321–14338.
- Anitha, P., & Kalpana, M. (2025). New RFM WL model for effective customer segmentation using machine learning. *Journal of Theoretical and Applied Information Technology*, *103*(4), 1420–1432.
- Antonius, A., & Fitrianah, D. (2025). Enhancing customer segmentation insights through an RFM and discount proportion model with clustering algorithms in e-commerce. *Journal of Computer Science and Information Technology*, *13*(1), 45–58.
- Aslantaş, G., Gençgül, M., Rumelli, M., Özsaraç, M., & Bakırlı, G. (2023). Customer segmentation
Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2009). *Database marketing: Analyzing and managing customers*. Springer Science & Business Media.

- Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking: An effective approach to customer segmentation. *Journal of King Saud University – Computer and Information Sciences*, *33*(10), 1252–1257.
- Cui, M. (2020). Introduction to the K-Means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, *1*(1), 157–164.
- Dol, S. M., & Jawandhiya, P. M. (2023). Classification technique and its combination with clustering: A review.
- Fakhar, I., Siddiqui, M. T., & Faizi, A. (2026). Effects of remittances on household welfare: An econometrics analysis from Kunar Province, Afghanistan. *Journal of Social Sciences and Humanities*, *3*(2), 451–467. https://doi.org/10.62810/jssh.v3i2.243
  Fatahi, S., & Rabiei, M. (2020). Users clustering based on search behavior analysis using the LRFM model and K-means algorithm. Journal of Big Data, 7(1), 1–18.
- Febriani, A., & Putri, S. A. (2020). Segmentasi konsumen berdasarkan model recency, frequency, monetary (RFM) menggunakan algoritma K-means. *Jurnal Systemic*, *6*(1), 23–30.
- Firdaus, U., & Utama, D. N. (2021). Development of bank's customer segmentation model based on RFM and K-means algorithm. *Indonesian Journal of Information Systems*, *4*(1), 45–56.
- Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science*, *25*(6), 718–739.
- Heldt, R., Silveira, C. S., & Luce, F. B. (2021). Predicting customer value per product: From RFM to RFM-D.
- *IEEE Transactions on Knowledge and Data Engineering*, *35*(2), 1120–1135.
  Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666.
- John, J. M., Shobayo, O., & Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the retail industry. *Data*, *8*(2), Article 34.
- *Journal of Business Research*, *129*, 444–453.
- Kamal, Y. (2026). *The symbiosis of marketing intelligence and customer experience: Building a data-driven strategy* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.6504998
- Kamal, Y., Rizvi, A. A., Ahmad, S., Siddiqui, M. T., Mishra, R. K., & Roy, A. (2027). AI-powered branding strategies' gravitational pull of innovation through the AI's impact on advertising *and* design. In *Human-AI cognitive collaboration for industrial evolution* (pp. 357–388). IGI Global. https://doi.org/10.4018/979-8-3373-7036-1.ch011 [Verify: Publication year 2027 is later than the current date; confirm the year.] [Verify: Book editor(s) not supplied.]
- Kamal, Y., Rizvi, A. A., Siddiqui, M. T., Ahmad, S., & Mishra, R. K. (2025). *Sustainable digital advertising: The paradox of carbon-neutral virality* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.6046194
- Kamal, Y., Rizvi, A. A., Srivastava, V., Ahmad, S., Kumar, S., & Siddiqui, M. T. (2026). Harnessing artificial intelligence for strategic retail promotions: A path to consumer-centric shopping. In *Driving global economic transformation through AI and machine learning* (pp. 363–398). IGI Global. [Verify: Book editor(s) not supplied.]
- Kamal, Y., & Siddiqui, M. T. (2025). *7 slot radiator grille and 70 years of mileage: Who drove off road on road?* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.5928514
- Kotler, P., Keller, K. L., & Chernev, A. (2022). *Marketing management* (16th ed.). Pearson Education.
- Kumar, N., Siddiqui, M. T., & Suhail, M. (2020). Theory of constraints: A review on its evolution and adoption. *The International Journal of Analytical and Experimental Modal Analysis*, *12*(9), 957–969. DOI: 18.0002.IJAEMA.2020.V12I9.200001.01568590054541 [Verify: The DOI as supplied (18.0002…) does not follow the DOI format, which always starts with “10.”; check it against the journal.]
- Kumar, V., Kamal, Y., Rizvi, A. A., Ahmad, S., Siddiqui, M. T., & Usman, M. (2025). *AI-native incrementality testing: Moving beyond traditional holdout groups in retail media* [Working paper]. SSRN. https://ssrn.com/abstract=5998474
- Kumar, V., & Reinartz, W. (2022). *Customer relationship management: Concept, strategy, and tools* (3rd ed.). Springer Nature.
- Kuraria, A., Jharbade, N., & Soni, M. (2018). Centroid selection process using WCSS and elbow method for K-Means clustering. *International Journal of Computer Sciences and Engineering*, *6*(5), 485–489.
- Li, M., Wang, Q., Shen, Y., & Zhu, T. Y. (2021). Customer relationship management analysis of outpatient department based on RFM model and K-means algorithm. *BMC Medical Informatics and Decision Making*, *21*(1), 1–11.
- Ma, J. (2022). E-commerce customer segmentation based on RFM model. In *Proceedings of International Conference on Big Data and Social Sciences* (pp. 112–119). Atlantis Press.

- Mansoori, M. V., Kumar, D., & Siddiqui, M. T. (2021). Role of artificial intelligence in sustainable consumption and production (SCP): A literature review perspective. *Shodh Sanchar Bulletin*, *11*(41), 283–288.
- Mishra, R. K., Ahmad, S., & Siddiqui, M. T. (2026). Voice commerce and conversational marketing: Understanding consumer decision paths and trust formation in audio-based AI interfaces. In *Encyclopedia of artificial intelligence in marketing* (pp. 1–16). Springer Nature Switzerland. [Verify: Book editor(s) not supplied.]
- Nasyuha, A. H., Zulham, Z., & Rusydi, I. (2022). Implementation of K-means algorithm in data analysis for customer segmentation. *Journal of Computer Science and Technology*, *2*(1), 14–22.
- Paul, L., & Ramanan, T. R. (2019). An RFM and CLV analysis for customer retention and customer relationship management. *International Journal of Recent Technology and Engineering*, *8*(2), 2200–2205.
- Perdana, S. A., Florentin, S. F., & Santoso, A. (2022). Analisis segmentasi pelanggan menggunakan model RFM dan K-means clustering. *Jurnal Teknoinfo*, *16*(2), 210–218.
- Perdhana, R. B., & Heikal, J. (2024). Enhancing customer segmentation in online transportation services using RFM and machine learning. *International Journal of Service Science*, *12*(1), 55–67.
- Pynadath, M. F., Rofin, T. M., & Thomas, S. (2023). Evolution *of* customer relationship management through machine learning: A systematic review. *Journal of Retailing and Consumer Services*, *72*, Article 103280.
- Reichheld, F. F., & Sasser, W. E. (1990). Zero defections: Quality comes to services. *Harvard Business Review*, *68*(5), 105–111.
- Rosalia, R., & Wibowo, A. (2026). Implementation of RFM model and K-Means clustering for customer segmentation in commercial service enterprise. *Journal of Information Systems and Applied Computer Science*, *15*(1), 101–115.
- Rungruang, C., Riyapan, P., Intarasit, A., Chuarkham, K., & Muangprathub, J. (2024). RFM model customer segmentation based on clustering techniques for online retail. *IEEE Access*, *12*, 18450–18462.
- Sabuncu, İ., Türkan, E., & Polat, H. (2020). Customer segmentation and profiling with RFM. *Turkish Journal of Marketing*, *5*(3), 201–215.
- Saha, L., Tripathy, H. K., Nayak, S. R., Bhoi, A. K., & Barsocchi, P. (2021). Amalgamation of customer segmentation using RFM model and DBSCAN clustering. *IEEE Access*, *9*, 132450–132462.
- Siddiqui, M. T., & Kamal, Y. (2026). Responsible automation in e-business: A framework for agentic AI integration and governance. *BIMTECH Business Perspectives*. https://doi.org/10.1177/25819542261479531 [Verify: Volume, issue and page numbers not supplied (may be online-first).]
- Siddiqui, M. T., Kamal, Y., Suhail, M., & Fakhar, I. (2026a). AI-enabled carbon accounting for MSMEs in emerging economies: Architecture, governance, and performance implications. In *Beyond the balance sheet: Sustainability and carbon accounting in green and blue economy* (pp. 6–18). LPU Publication House. [Verify: Book editor(s) not supplied.]
- Siddiqui, M. T., Kumar, S., Bharti, S., Ahmad, Y., & Suhail, M. (2026b). Fairness, accountability, *and* transparency in financial AI: Addressing bias through responsible regulation *and* auditable design. *Minnesota Journal of Business Law and Entrepreneurship*, (2), 1547–1557. https://doi.org/10.52783/mjble.382 [Verify: Volume number not supplied (only issue 2).]
- Siddiqui, M. T., Mansoori, M. V., Siddiqui, M. A., & Yadav, A. (2025). AI-enabled pedagogy: Advancing education through innovative teaching tools and the AI-TEACH model. *Journal of Informatics Education and Research*, *5*(1), 2526–2539. https://doi.org/10.52783/jier.v5i1.2261
- Siddiqui, U. A., Mehdi, Z., & Siddiqui, M. T. (2024). An empirical study exploring the determinants of consumer behavior in e-retailing. *Communication and Management*, *9*(10), 138–153. https://doi.org/10.36896/CMJ2024.V9I10.24.21730
- Silamantha, W. A., & Hadiono, K. (2024). Analisis RFM dan K-Means Clustering untuk segmentasi pelanggan toko retail. *Jurnal Teknologi Informasi dan Komputer*, *10*(1), 77–88.
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727.
- Smaili, M. Y., & Hachimi, H. (2023). New RFM-D classification model for improving customer analysis using machine learning algorithms. *Decision Science Letters*, *12*(3), 485–496.
- Sun, Y., Liu, H., & Gao, Y. (2023). Research on customer lifetime value based on machine learning and extended RFM model. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(2), 890–905.
- Syahra, Y., Fadlil, A., & Yuliansyah, H. (2025). Customer segmentation using RFM and K-Means clustering algorithm for small retail enterprise. *International Journal of Computer and Information System*, *6*(1), 12–21.

- Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means clustering approach for intelligent customer segmentation using customer purchase behavior. *Sustainability*, *14*(12), Article 7243.
- Taqwim, A., & Suprobo, A. (2019). Application of RFM model for customer segmentation in e-commerce using K-means. *Journal of Information Technology and Computer Science*, *4*(2), 115–124.
- Tyagi, S., & Gupta, P. (2024). Customer segmentation in e-commerce using machine learning techniques: A comparative study. *Journal of Retailing and Consumer Services*, *78*, Article 103720.
- Umisha, E., & EEL Team. (2024). Data-driven customer segmentation and churn prediction using RFM and unsupervised clustering. *Journal of Electronic Commerce Research*, *25*(3), 178–194.
- using RFM model and K-means algorithm: An application in the e-commerce sector. *Journal of Information Systems and Management Research*, *5*(2), 88–97.
- Vouros, A., Langdell, S., Croucher, M., & Vasilaki, E. (2021). An empirical comparison between stochastic and mini-batch K-means algorithms. *Neurocomputing*, *452*, 112–124.
- Wahyuningrum, T., Khomsah, S., Suyanto, S., Meliana, S., Yunanto, P. E., & Al Maki, W. F. (2021). Customer segmentation using RFM model and K-means clustering on e-commerce dataset. *Journal of Physics: Conference Series*, *1869*(1), Article 012085.
- Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Springer Science & Business Media.
- Wulansari, S., & Heikal, J. (2024). Analysis of customer segmentation in the top three most visited e-commerce platforms using RFM and K-means. *Journal of Marketing Analytics*, *12*(2), 145–158.
- Yadav, S., Ahmad, I., Kamal, Y., Rizvi, A. A., & Siddiqui, M. T. (2026). *Promoting cultural sensitivity in bleisure tourism: Strategies for sustainable engagement* [Working paper]. SSRN. https://doi.org/*10*.2139/ssrn.6208798
- Yulisasih, B. N., Herman, H., Sunardi, S., & Yuliansyah, H. (2024). Evaluation of K-Means and Fuzzy C-Means clustering algorithms for customer segmentation based on RFM model. *Journal of Applied Data Sciences*, *5*(2), 89–101.
- Zhu, N. (2023). Research on customer relationship segmentation of apparel retail industry through RFM model and K-means algorithm. *Highlights in Business, Economics and Management*, *10*, 234–242.