**Jurnal EMT KITA**
*Indonesian Journal for the Economics, Management and Technology*
**Print ISSN: 2579-7972; Online ISSN: 2549-6204**

**Volume 9 (4) October 2025, 1850-1859**
Available *online* at
http://journal.lembagakita.org/index.php/emt

Article History: Received: 14 July 2025, Revision: 20 July 2025, Accepted: 20 August 2025, Available *Online*: 1 October 2025.

DOI: https://doi.org/10.35870/emt.v9i4.5375

# E-Commerce Customer Segmentation Based on RFM Analysis Using DBSCAN Algorithm to Improve Marketing Strategy

**Ferry Irawan** [1], **Fitroh Amaluddin** [2*], **Aris Wijayanti** [3]
[1,2*,3] Universitas PGRI Ronggolawe, Indonesia.

*Corresponding Email* : amfitroh@gmail.com [2*]

**Abstrak.** *Segmentasi pelanggan merupakan langkah penting dalam merancang strategi pemasaran yang efektif, khususnya dalam bisnis e-commerce yang memiliki basis pelanggan luas dengan karakteristik perilaku belanja yang beragam. Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan pendekatan RFM (Recency, Frequency, dan Monetary) guna mengidentifikasi perilaku pelanggan dan mendukung pengambilan keputusan pemasaran yang lebih tepat sasaran. Algoritma yang digunakan dalam proses clustering adalah DBSCAN (Density-Based Spatial Clustering of Applications with Noise) karena kemampuannya dalam menemukan klaster dengan bentuk arbitrer dan mendeteksi outlier atau pelanggan tidak aktif. Proses penelitian dimulai dengan pemilihan dataset, dilanjutkan dengan preprocessing data melalui tahap pembersihan, reduksi, transformasi RFM, dan normalisasi menggunakan metode Min-Max. Penelitian ini membuktikan bahwa kombinasi pendekatan RFM dan algoritma DBSCAN efektif dalam melakukan segmentasi pelanggan e-commerce. Hasil segmentasi ini dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih personal, seperti retargeting pelanggan aktif dan reaktivasi pelanggan tidak aktif, sehingga dapat meningkatkan efisiensi pemasaran dan loyalitas pelanggan.*

**Kata kunci:** *Segmentasi Pelanggan; E-Commerce; Strategi Pemasaran.*

**Abstract.** *Customer segmentation is an important step in designing an effective marketing strategy, especially in e-commerce businesses that have a large customer base with diverse shopping behavior characteristics. This study aims to segment customers based on the RFM (Recency, Frequency, and Monetary) approach to identify customer behavior and support more targeted marketing decision-making. The algorithm used in the clustering process is DBSCAN (Density-Based Spatial Clustering of Applications with Noise) because of its ability to find clusters with arbitrary shapes and detect outliers or inactive customers. The research process began with the selection of the dataset, followed by data preprocessing through the stages of cleaning, reduction, RFM transformation, and normalization using the Min-Max method. This study proves that the combination of the RFM approach and the DBSCAN algorithm is effective in segmenting e-commerce customers. The results of this segmentation can be used to develop more personalized marketing strategies, such as active customer retargeting and inactive customer reactivation, so that it can increase marketing efficiency and customer loyalty.*

**Keywords:** *Customer Segmentation; E-Commerce; Marketing Strategy.*

## Introduction

Grounded in the rapid growth and intensifying competition of the e-commerce sector, personalization and customer segmentation have become foundational for balancing acquisition needs with retention goals while optimizing marketing budgets. Empirical evidence shows that personalization enhances customer experience and satisfaction and, in turn, stimulates repeat purchase intentions (Gomes & Meisen, 2023; Madu & Manggu, 2024; Wisnel *et al.*, 2022). Advances in artificial intelligence and recommender systems further expand platforms' capacity to infer preferences and deliver relevant offers in real time (Erawati *et al.*, 2023; Riswan *et al.*, 2024). Consequently, precise segmentation affects not only user experience but also campaign effectiveness and long-term profitability. Concurrently, firms face high customer acquisition costs and increasingly heterogeneous shopping behaviors, especially in the post-pandemic period as online channels accelerated adoption (Kaya *et al.*, 2024; Purnama & Bangun, 2024; Wang, 2024).

A stronger focus on retention is warranted because of its close linkage to firm performance and valuation through customer lifetime value (CLV) (Datta *et al.*, 2018; Livne *et al.*, 2011; Yasfi & Pardede, 2023). Improvements in service quality, effective communication, and user interface/user experience (UI/UX) reliably foster satisfaction and loyalty (Duffour *et al.*, 2022; Halim & Berlianto, 2024; Rakhman *et al.*, 2022; Sutisna & Sutrisna, 2023). These complexities motivate data-driven segmentation approaches capable of capturing variation in customer behavior with managerial precision (Abbu & Gopalakrishna, 2022; Nitzan & Libai, 2011). Within behavioral segmentation, the Recency Frequency Monetary (RFM) framework is a well-established proxy for customers' temporal proximity, transactional intensity, and value, making it directly relevant to marketing prioritization and CLV strengthening (Christy *et al.*, 2021; Gustriansyah *et al.*, 2020; Londhe & Palwe, 2022). RFM facilitates grouping customers with similar profiles and translating these profiles into targeted retention, cross-sell, and win-back programs (Abe, 2016; Kabasakal, 2020; Samidi *et al.*, 2023). In practice, however, success with RFM depends not only on metric construction but also on clustering techniques that can accommodate transactional data that are typically skewed and outlier-prone (Abednego *et al.*, 2023). Prior studies have combined RFM with multiple clustering algorithms from K-Means and K-Medoids to Fuzzy C-Means and density-based methods to extract actionable segments (Brahmana *et al.*, 2020; Fadhillah *et al.*, 2024; Prasetyo *et al.*, 2020; Singh *et al.*, 2023). Yet partitioning methods such as K-Means require specifying the number of clusters a priori, are sensitive to feature scaling and outliers, and often assume roughly spherical cluster shapes assumptions frequently violated by RFM data in e-commerce (Brahmana *et al.*, 2020; Wong *et al.*, 2024). These limitations justify algorithms that are more robust to noise and can follow irregular cluster contours. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) stands out as a compelling alternative because it is density-based, identifies noise/outliers, does not require pre-specifying the number of clusters, and flexibly discovers arbitrarily shaped clusters properties well aligned with RFM data characteristics (Brahmana *et al.*, 2020; Kusuma & Sudiartha, 2020).

Recent empirical evidence underscores DBSCAN's effectiveness for customer segmentation across sectors, including e-commerce, and its ability to reveal non-linear patterns often missed by partitioning algorithms (Monalisa *et al.*, 2023; Wong *et al.*, 2024). On this basis, integrating RFM with DBSCAN is expected to yield stable, interpretable segments ready to be mapped to differentiated marketing tactics. Despite this promise, notable scholarly and practical gaps remain. Many studies stop at cluster formation without explicitly linking segments to concrete marketing performance indicators (e.g., click-through rate/CTR, conversion rate/CVR, average order value/AOV) or to operational playbooks for commercial teams; furthermore, evidence set in the Indonesian e-commerce context that ties RFM+DBSCAN to metrically grounded marketing decisions is still relatively limited (Fadhillah *et al.*, 2024; Gomes & Meisen, 2023; Kusuma & Sudiartha, 2020; Monalisa *et al.*, 2023; Prasetyo *et al.*, 2020). Addressing these gaps is crucial given post-pandemic heterogeneity in behavior and the rising need for cross-channel personalization. This study applies DBSCAN to RFM features to construct robust e-commerce customer segments, map each segment to executable marketing strategies—covering value propositions, channels, and communication frequency and evaluate impact using a blend of cluster-validity metrics and business indicators. The study also documents preprocessing steps (robust transformations/scaling and outlier handling), rationale for parameterization (ε and minPts), and a concise comparison with partitioning baselines such as K-Means to assess trade-offs in accuracy, interpretability, and ease of deployment (Brahmana *et al.*, 2020; Singh *et al.*, 2023; Wong *et al.*, 2024). Overall, the article's contributions are both conceptual and practical: (1) an integrative RFM+DBSCAN framework tailored to transactional data characteristics; (2) explicit linkage between cluster outputs and measurable, segment-specific marketing recommendations; and (3) evidence relevant to decision-makers in the Indonesian e-commerce setting. The remainder of the paper is organized as follows: the methods section details RFM construction, preprocessing, and DBSCAN configuration; the results present segment profiles and evaluation; the discussion elaborates strategic implications and limitations; and the conclusion summarizes findings and outlines directions for future research (Christy *et al.*, 2021; Gomes & Meisen, 2023; Wong *et al.*, 2024).

## Research Methodology

### Dataset Selection

This study employs the publicly available Online Retail dataset from Kaggle. The data comprise transaction logs from a UK-based online retailer over a defined observation window. In total, the dataset contains 541,911 rows and eight core attributes InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country each capturing a distinct facet of the online sales process. The dataset was chosen because its openness and breadth provide a rich representation of purchasing behavior and make it well suited for segmentation using the RFM (Recency, Frequency, Monetary) framework.

### Preprocessing

Preprocessing ensures the analytical dataset is clean, relevant, and properly structured before modeling. Two major steps were undertaken: data cleaning and data reduction. Data cleaning involved removing invalid observations. Specifically, records with missing values in key fields (e.g., CustomerID, InvoiceNo) were discarded; transactions with non-positive Quantity or UnitPrice likely returns or data entry errors were excluded; and exact duplicate rows were dropped. After cleaning, the dataset was reduced from 541,911 to 22,192 valid records. Data reduction then retained only attributes essential to RFM analysis. From the original eight fields, five were preserved because they map directly to RFM construction: CustomerID (customer identifier), InvoiceNo (for counting transaction frequency), InvoiceDate (to determine the most recent purchase), and Quantity with UnitPrice (to compute purchase value). Variables such as StockCode, Country, and Description were omitted because they do not contribute to RFM modeling. This step lowers complexity and focuses the analysis on information that is strictly necessary.

### Transformation

Next, RFM features were derived at the customer level. Recency was calculated as the number of days between a reference date (the latest date in the dataset) and each customer's most recent transaction. Frequency was measured as the count of unique invoices (InvoiceNo) per customer. Monetary represented the total spending per customer, computed as the sum of (Quantity × UnitPrice). Because Monetary typically exhibits a larger scale than Recency and Frequency, min–max scaling was applied to place features on comparable ranges, mitigating scale bias and improving clustering performance.

### Clustering

With transformed and normalized features, customers were segmented using DBSCAN

(Density-Based Spatial Clustering of Applications with Noise). DBSCAN was selected because it can recover clusters of arbitrary shape and explicitly label noisy or outlying points capabilities that are advantageous for transactional data. The algorithm's key parameters, epsilon (ε) and the minimum number of points (MinPts), were tuned using inspection of the k-distance (k-nearest neighbor distance) plot and parameter experimentation to obtain well-separated, meaningful clusters.

### Evaluation

Clustering quality was assessed using the Silhouette score, which quantifies how well each point fits within its assigned cluster relative to other clusters, taking values from −1 to 1; higher values indicate better separation and cohesion. In addition to the numerical metric, scatter-plot visualizations were produced to inspect the spatial distribution of clusters. These plots support a deeper interpretation of the behavioral profiles associated with each segment and facilitate business-oriented insights.

## Results and Discussion

### Results

### Results Prepocessing

The preprocessing stage begins with data cleaning, which aims to improve the quality of the data and ensure that only valid data is used in the analysis. The initial dataset consisted of 541,911 rows of data. After deleting empty values, transactions with a Quantity or UnitPrice of 0 or negative, and duplicate data, the amount of data is significantly reduced to 22,192 lines of net data. Next, the data reducing process is carried out, which is simplifying attributes to only five columns that are relevant to the RFM analysis, namely CustomerID, InvoiceNo, InvoiceDate, Quantity, and UnitPrice.

Table 1. Sample Data Before Cleaning

| InvoiceNo | CustomerID | Stock Date | … | Description |
|---|---|---|---|---|
| 537236 | 16858 | 22073 | … | Red Retrospot Storage Jar |
| 537237 | | 15036 | … | Assorted Colours Silk Fan |
| 580724 | | 23389 | … | Spaceboy Mini Backpack |
| … | … | … | … | … |
| 581587 | 12680 | 22138 | … | Baking Set 9 Piece Retrospot |

Table 2. Sample Data After Cleaning

| InvoiceNo | CustomerID | Stock Date | … | Description |
|---|---|---|---|---|
| 537236 | 16858 | 22073 | … | Red Retrospot  Storage Jar |
| 537236 | 16858 | 21216 | … | Set 3 Retrospot Tea,Coffee,Sugar |
| 537236 | 16858 | 21527 | … | Red Retrospot Traditional Teapot |
| … | … | … | … | … |
| 581587 | 12680 | 22138 | … | Baking Set 9 Piece Retrospot |

Table 3. Data Reduction

| CustomerID | Quantity | Unit Price | InvoiceNo | InvoiceDate |
|---|---|---|---|---|
| 17850 | 6 | 2.55 | 536365 | 01/01/2020 |
| 17850 | 8 | 3.39 | 536365 | 01/01/2020 |
| 17850 | 8 | 2.75 | 536365 | 01/01/2020 |
| … | … | … | … | … |
| 12680 | 3 | 4.95 | 581587 | 31/12/2020 |

### RFM Data Transformation Results

Data transformation is carried out using the RFM (Recency, Frequency, Monetary) model approach to evaluate customer behavior based on three main aspects, namely how recently the customer made the last transaction (Recency), how often the customer made the transaction (Frequency), and how much the total value of the transaction was made by the customer (Monetary). The initial step of this transformation is to calculate the Recency value for each customer, i.e. by measuring the day difference between the customer's last transaction date and the reference date (the last day in the dataset). Then, Frequency is calculated based on the number of unique transactions made by customers. Finally, Monetary is calculated by summing the total spending of each customer obtained from the multiplication between the number of goods (Quantity) and the unit price (UnitPrice).

Table 4. Data Transformation

| Recency | Frequency | Monetery |
|---|---|---|
| 2 | 7 | 425492 |
| 75 | 4 | 147324 |
| 19 | 1 | 141366 |
| … | … | … |
| 36 | 11 | 120179 |

After obtaining the RFM value, the next stage is data normalization. This normalization process is necessary because the scale between the three attributes has a very different range of values, especially Monetary which has a much larger number than Recency and Frequency. Therefore, Min-Max Normalization is performed to convert all RFM values to a range between 0 to 1. This is important so that clustering algorithms such as DBSCAN can provide more accurate results and are not biased towards attributes at the largest scale.

Table 5. Data Normalisasi

| Recency | Frequency | Monetery |
|---|---|---|
| 0.0026 | 0.0242 | 0.0314 |
| 0.1983 | 0.0121 | 0.0212 |
| 0.0482 | 0.0000 | 0.0210 |
| … | … | … |
| 0.0938 | 0.0404 | 0.0202 |

**Clustering Results with DBSCAN**

Clustering is a data segmentation technique used to group customers into several groups based on the similarity of their shopping behavior characteristics. In this study, the algorithm used for the clustering process is DBSCAN (Density-Based Spatial Clustering of Applications with Noise). The selection of DBSCAN was motivated by its ability to identify cluster patterns with irregular shapes (arbitrary), as well as its advantages in handling data that contains noise or outliers. This is especially relevant in the context of customer transaction data that is often not uniformly distributed and contains anomalies. In order for DBSCAN to work optimally, two important parameters need to be determined, namely epsilon (ε) and min_samples. Epsilon is the maximum distance between two points so that it can be considered a neighbor in a cluster. If the epsilon value is too small, then the data that has similarities may not be in the same cluster. Conversely, if it is too large, then the cluster can become too large and mix data that is actually different. Meanwhile, min_samples is the minimum number of points needed to form a solid cluster (core point). A min_samples value that is too small will lead to many small clusters that are meaningless, whereas if it is too large, then many points will be categorized as noise because they are not qualified to form clusters. To get the most optimal combination of parameters, two analysis approaches were carried out, namely the K-Distance Graph and grid search with evaluation using Silhouette Score. First, the K-Distance Graph is used to estimate the value of epsilon. This graph is created by calculating the closest distance from each point to the nearest neighbor k (with k = min_samples - 1), and then plotting the results sequentially. The elbow point on this graph is considered the optimal epsilon value because it shows a sharp change in the slope of the graph that marks the difference between the solid cluster and the distribution of data outside the cluster. Second, a grid search was conducted to explore several combinations of epsilon and min_samples values, which were then evaluated using the Silhouette Score. This metric measures how well the data has been clustered, with a value close to +1 indicating that the data point is in the right cluster, while a value close to -1 indicates a possible misclustering.

Table 6. Optimal Parameter Values

| Epsilon | Min Samples |
|---|---|
| 0.1026 | 3 |

Through this process, the best combination of parameters was obtained, namely epsilon of 0.1026 and min_samples of 3. With these parameters, the DBSCAN algorithm manages to form a sharper and more meaningful cluster structure, and can identify outlier points more accurately. This makes customer segmentation more informative and useful for strategic decision-making. The visualization of clustering results was carried out using a 3-dimensional scatter plot based on normalized Recency, Frequency, and Monetary attributes. Through this visualization, the distribution pattern of customers in each cluster can be observed more clearly, including how close it is between customers in a cluster, the density of each cluster, and the presence of customers who are included in the noise or inactive category.

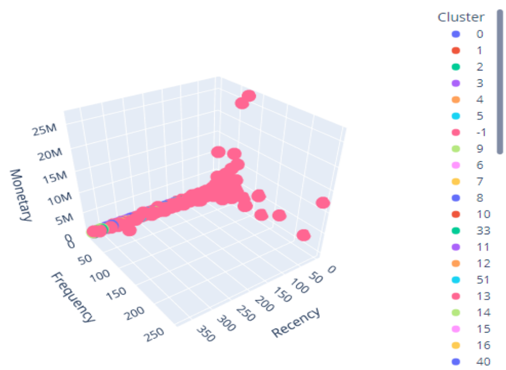


Figure 1. Visualization of Clustering Results

Evaluation of the clustering results showed that the Silhouette Score obtained was quite satisfactory, which indicated that the formation of the cluster was in accordance with the data structure. The clusters that form are quite separate and dense, while the noise points are scattered around the boundaries between the clusters. These results show that DBSCAN is effective in capturing segmentation of customer behavior. Overall, customer segmentation obtained from the results of this clustering can provide strategic benefits for the company. For example, clusters with high Frequency and Monetary values can be targeted with loyalty programs or exclusive offers, while clusters with high Recency values can be targeted for customer reactivation campaigns. Meanwhile, customers who fall under the noise category need to be further analyzed to determine if they are still worth following up on or need to be eliminated from the marketing target. Thus, the implementation of DBSCAN on customer data not only provides an in-depth understanding of consumer behavior, but also supports the planning of marketing strategies that are more targeted and efficient.

**Discussion**

The data curation results indicate that rigorous preprocessing was essential to obtain a behaviorally credible sample for segmentation. From 541,911 raw records, the removal of missing identifiers, non-positive quantities or prices (likely returns or entry errors), and exact duplicates yielded 22,192 valid transactions. This stringent filtering minimizes measurement noise that can blur behavioral signals in RFM construction and density estimation, a concern that is particularly acute in e-commerce where purchase logs are heterogeneous and error-prone. By concentrating the analysis on valid, behavior-bearing events, the study aligns with calls for data-driven personalization and segmentation capable of capturing genuine variation in customer behavior an imperative linked to user experience, loyalty, and campaign efficiency (Abbu & Gopalakrishna, 2022; Madu & Manggu, 2024; Rakhman *et al*., 2022; Sutisna & Sutrisna, 2023; Wisnel *et al*., 2022). The transformation of cleaned transactions into Recency, Frequency, and Monetary (RFM) features provides an interpretable bridge between transactional traces and marketing priorities. Recency approximates temporal engagement, Frequency captures interaction intensity, and Monetary reflects value contribution; together they offer a compact, managerially meaningful basis for targeting (Christy *et al*., 2021; Gustriansyah *et al*., 2020; Londhe & Palwe, 2022). Normalizing RFM via min–max scaling is appropriate given the scale dominance and skew typically exhibited by spending, reducing the risk that any single dimension overwhelms clustering. These design choices are consistent with prior work that links RFM profiles to actionable programs (retention, cross-sell, win-back) and to customer lifetime value (CLV), thereby tying segmentation outcomes to long-run profitability (Abe, 2016; Abednego *et al*., 2023; Kabasakal, 2020; Livne *et al*., 2011).

Against this representation, DBSCAN was an apt modeling choice. Unlike partitioning methods such as K-Means that assume roughly spherical clusters and require the number of clusters a priori, DBSCAN discovers arbitrarily shaped clusters and explicitly labels noise capabilities that fit skewed, outlier-laden RFM data (Brahmana *et al*., 2020; Wong *et al*., 2024). Parameterization via the k-distance elbow and a grid search yielded $\varepsilon = 0.1026$ and min_samples = 3, which produced sharp, behaviorally coherent groupings in three-dimensional RFM space. This approach accords with findings that density-based methods perform well for customer segmentation, including in e-commerce settings, particularly when distributions are uneven or when outliers are informative rather than nuisance variance (Kusuma & Sudiartha, 2020; Monalisa *et al*., 2023; Singh *et al*., 2023). At the same time, the use of a relatively small min_samples trades off sensitivity to fine-grained substructures against a higher risk of spurious micro-clusters an issue mitigated here by joint inspection of the k-distance plot and external validity considerations. Evaluation with the Silhouette score indicated satisfactory separation and cohesion, with outliers concentrated near cluster boundaries. This pattern mirrors evidence that DBSCAN can yield well-differentiated behavioral segments when densities are meaningfully distinct, while also surfacing noise that may correspond to infrequent or anomalous shoppers (Monalisa *et al*., 2023; Wong *et al*., 2024). Where some studies have reported competitive performance for K-Means, K-Medoids, or Fuzzy C-Means on RFM (Brahmana *et al*., 2020; Fadhillah *et al*., 2024; Prasetyo *et al*., 2020), those methods tend to be more scale- and centroid-sensitive, and they lack an intrinsic noise model. The present results therefore complement, rather than contradict, prior work: when the analyst anticipates non-spherical structures and meaningful outliers, density-based clustering is theoretically and empirically justified; when clusters are compact and well balanced, partitioning methods may suffice.

Managerially, the discovered segments map naturally to differential strategies. High-Frequency/Monetary clusters are candidates for loyalty and exclusivity programs, while high-Recency/low-engagement groups suit reactivation offers; the noise set merits diagnostic review to distinguish nascent customers from transaction anomalies. These prescriptions are consistent with research linking personalization, service quality, and seamless experiences to higher satisfaction, retention, and CLV (Gomes & Meisen, 2023; Madu & Manggu, 2024; Nitzan & Libai, 2011; Wisnel *et al*., 2022). They also align with advances in AI-driven recommendation that can operationalize segment-specific propositions across channels (Erawati *et al*., 2023; Riswan *et al*., 2024). Two caveats warrant future work: first, the substantial reduction during cleaning, while defensible, may constrain generalizability across categories and seasons; second, comparing DBSCAN with tuned partitioning baselines on multiple validity indices and business KPIs (e.g., CTR, CVR, AOV, and CLV lift) would clarify trade-offs under varying density regimes. Even with these limitations, the evidence supports density-based

RFM segmentation as a practical pathway to scalable personalization in heterogeneous e-commerce markets.

## Conclusion

Based on the results of the research that has been conducted, it can be concluded that the DBSCAN algorithm is able to group customers based on their shopping behavior using the RFM (Recency, Frequency, Monetary) approach effectively. The data normalization process using the Min-Max method is carried out to equalize the scale between attributes, so that the grouping results become more accurate. After searching for optimal parameters using the K-Distance Graph and evaluation through Silhouette Score, an epsilon value of 0.1026 and a min_samples of 3 were obtained as the best parameters. With this configuration, DBSCAN manages to form several clusters that reflect different customer behavior patterns as well as identify a certain amount of data as noise that can be considered an inactive customer or outlier. The results of the visualization and evaluation show that the separation between clusters is quite clear, and the resulting quality of the clusters can support decision-making in more targeted marketing strategies, such as retargeting of potential customers or reactivation of inactive customers.

As a suggestion, for further research, it is recommended to compare with other clustering algorithms such as K-Means or Agglomerative Clustering to find out which approach produces the most optimal segmentation for the customer data used. Additionally, it would be better to use customer data with richer attributes, such as product preferences, purchase time, and purchase channels (online/offline), to generate more relevant segmentation. This research can also be further developed by associating the results of clustering with digital marketing strategies or the development of a recommendation system based on customer clusters.

## References


Abbu, H., & Gopalakrishna, P. (2022). Digital transformation powered by big data analytics: The case of retail grocery business. *Hawaii International Conference on System Sciences*, 231. https://doi.org/10.24251/hicss.2022.231.

Abe, M. (2016). Deriving customer lifetime value from RFM measures: Insights into customer retention and acquisition. In *Customer Relationship Management* (pp. 127–160). MIT Press. https://doi.org/10.7551/mitpress/9471.003.0007.

Abednego, L., Nugraheni, C. E., & Salsabina, A. (2023). Customer segmentation: Transformation from data to marketing strategy. *Conference Series, 4*(1), 139–152. https://doi.org/10.34306/conferenceseries.v4i1.645.

Brahmana, R. W., Mohammed, F. A., & Chairuang, K. (2020). Customer segmentation based on RFM model using K-Means, K-Medoids, and DBSCAN methods. *LONTAR KOMPUTER*. https://doi.org/10.24843/LKJITI.2020.v11.i01.p04.

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking—An effective approach to customer segmentation. *Journal of King Saus University - Computer Science and Information Sciences*, 1251–1257.

Datta, P., Fraser, P., & Lebcir, M. (2018). An investigation into the relationship between customer relationship marketing and customer retention: Superstore retailing context in Bangladesh. *Journal of Business & Retail Management Research, 13*(02). https://doi.org/10.24052/jbrmr/v13is02/art-03.

Duffour, K. A., Gbolonyo, P. K., & Tandoh, I. (2022). Service quality in shipping industry of Ghana—Determinants and impact on customer satisfaction and customer retention. *International Journal of Applied Research in Business and Management, 3*(2), 112–134. https://doi.org/10.51137/ijarbm.2022.3.2.8.

Erawati, S. H., Amalia, N. S., Mauludin, H., & Liana, Y. (2023). Peran brand image sebagai moderator celebrity endorser pada keputusan pembelian. *Akuntansi Bisnis & Manajemen (ABM), 30*(2). https://doi.org/10.35606/jabm.v30i2.1276.

Fadhillah, M. F., Suyoso, A. L. A., & Puspitasari, I. (2024). Segmentasi pelanggan dengan algoritma clustering berdasarkan atribut Recency, Frequency dan Monetary (RFM): *Customer Segmentation with Clustering Algorithm Based on Recency, Frequency, and Monetary (RFM) Attributes*. *MALCOM, 5*(1), 48–56.

Gomes, M. A., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and E-Business Management, 21*, 527–570.

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on K-Means. *Indonesian Journal of Electrical Engineering and Computer Science, 18*(1), 470. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477.

Halim, I., & Berlianto, M. P. (2024). The effect of service quality dimensions, perceived value, customer satisfaction, and brand love toward customer retention on First Media's subscribers. *Widya Cipta - Jurnal Sekretari Dan Manajemen, 8*(2), 139–152. https://doi.org/10.31294/widyacipta.v8i2.20052.

Kabasakal, İ. (2020). Customer segmentation based on Recency Frequency Monetary model: A case study in e-retailing. *Bilişim Teknolojileri Dergisi, 13*(1), 47–56.

Kaya, C., Gemalmaz, S., Üzümcüoğlu, Y., Kartal, M., & Ceylan-Batur, S. (2024). How Covid-19 changed consumer behavior trends in Türkiye? *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 26*(2), 915–945. https://doi.org/10.16953/deusosbil.1375618.

Kusuma, A. A. N. W., & Sudiartha, I. M. (2020). Segmentasi pelanggan menggunakan RFM dan algoritma DBSCAN untuk mendukung strategi pemasaran. *Jurnal Ilmiah Merpati, 8*(3), 143–152.

Livne, G., Simpson, A., & Talmor, E. (2011). Do customer acquisition cost, retention, and usage matter to firm performance and valuation? *Journal of Business Finance & Accounting, 38*(3–4), 334–363. https://doi.org/10.1111/j.1468-5957.2010.02229.x.

Londhe, S. T., & Palwe, S. (2022). Customer-centric sales forecasting model: RFM-ARIMA approach. *Business Systems Research Journal, 13*(1), 35–45. https://doi.org/10.2478/bsrj-2022-0003.

Madu, M. I., & Manggu, B. (2024). Pengaruh E-Service Quality, brand image, kualitas produk terhadap keputusan pembelian ulang di e-commerce di Kabupaten Bengkayang. *Conference, 79–91*. https://doi.org/10.31603/conference.12042.

Monalisa, S., Juniarti, Y., Saputra, E., Muttakin, F., & Ahsyar, T. K. (2023). Customer segmentation with RFM models and demographic variable using DBSCAN algorithm. *Telkomnika (Telecommunication Computing Electronics and Control), 21*(4), 742–749.

Nitzan, I., & Libai, B. (2011). Social effects on customer retention. *Journal of Marketing, 75*(6), 24–38. https://doi.org/10.1509/jm.10.0209.

Prasetyo, S., Mustafid, M., & Hakim, A. (2020). Penerapan fuzzy c-means kluster untuk segmentasi pelanggan e-commerce dengan metode Recency Frequency Monetary (RFM). *Jurnal Gaussian, 9*(4). https://doi.org/10.14710/j.gauss.v9i4.29445.

Purnama, S., & Bangun, C. S. (2024). Strategic management insights into housewives consumptive shopping behavior in the post COVID-19 landscape. *Aptisi Transactions on Management (Atm), 8*(1), 71–79. https://doi.org/10.33050/atm.v8i1.2209.

Rakhman, A., Wirianto, H., Bunfa, L., & Rifiyanti, H. (2022). Faktor—Faktor e-commerce yang memengaruhi tingkat konsumerisme di Jakarta. *Jurnal Ilmiah Hospitality, 11*(1), 245–258. https://doi.org/10.47492/jih.v11i1.1608.

Riswan, D., Putra, H. E. R., & Saputra, R. N. (2024). Pengembangan sistem rekomendasi berbasis kecerdasan buatan untuk meningkatkan pengalaman pengguna di platform e-commerce. *Juktisi, 2*(3), 572–580. https://doi.org/10.62712/juktisi.v2i3.145.

Samidi, Suladi, R. Y., & Kusumaningsih, D. (2023). Comparison of the RFM model's actual value and score value for clustering. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi), 7*(6), 1430–1438. https://doi.org/10.29207/resti.v7i6.5416.

Singh, R., Patira, R., Singh, A., Choudhary, P., & Gupta, B. K. (2023). Customer segmentation on e-commerce data using clustering and RFM analysis. *Data Science and Analytics: Proceedings of the International Conference on Data Science and Analytics 2023*, 345–356.

Sutisna, N., & Sutrisna. (2023). Implementasikan sistem informasi dalam mendukung perilaku pembelian terhadap keputusan pembelian e-commerce. *Jurnal Mentari Manajemen Pendidikan Dan Teknologi Informasi, 2*(1), 20–30. https://doi.org/10.33050/mentari.v2i1.343.

Wang, X.-Y. (2024). Digital marketing strategies and consumer engagement: Unveiling TikTok's e-commerce dynamics. *Advances in Economics Management and Political Sciences, 103*(1), 183–190. https://doi.org/10.54254/2754-1169/103/20242445.

Wisnel, W., Wirdianto, E., & Cantika, T. (2022). Analysis of the effect of e-service quality on e-commerce customer satisfaction and loyalty. *Motivection Journal of Mechanical Electrical and Industrial Engineering, 4*(3), 209–222. https://doi.org/10.46574/motivection.v4i3.145.

Wong, C.-G., Tong, G.-K., & Haw, S.-C. (2024). Exploring customer segmentation in e-commerce using RFM analysis with clustering techniques. *Journal of Telecommunications and the Digital Economy, 12*(3), 97–125.

Yasfi, S. M., & Pardede, R. (2023). Influence of e-service quality towards customer retention and word of mouth (WOM) mediated by customer satisfaction. *Jurnal Indonesia Sosial Sains, 4*(11), 1168–1180. https://doi.org/10.59141/jiss.v4i11.927.