DOI:10.14421/ijid.2024.4841 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 13, No. 2, December 2024, Pp. 473-484

Analyzing Customer Loyalty Levels through Segmentation in Aesthetic Clinics Using K-Means and RFAM

Sinarring Azi Laga*
Department of Informatics
Universitas Hayam Wuruk Perbanas
Surabaya, Indonesia
sinarring.laga@perbanas.ac.id

Deny Hermansyah
Department of Informatics
Universitas Hayam Wuruk Perbanas
Surabaya, Indonesia
deny.hermansyah@hayamwuruk.ac.id

Chitra Laksmi Rithmaya
Department of Banking Financial Management
Universitas Hayam Wuruk Perbanas
Surabaya, Indonesia
citra@perbanas.ac.id

Muhammad Zainuddin
Department of Informatics
Universitas Hayam Wuruk Perbanas
Surabaya, Indonesia
202202011008@mhs.hayamwuruk.ac.id

Geo Ardana Ihsan Purnama Aji
Department of Informatics
Universitas Hayam Wuruk Perbanas
Surabaya, Indonesia
202202011005@mhs.hayamwuruk.ac.id

Iqbal Ramadhani Mukhlis
Department of Informatics
Universitas Hayam Wuruk Perbanas
Surabaya, Indonesia
iqbal.ramadhani@perbanas.ac.id

Article History
Received October 23rd, 2024
Revised December 2nd, 2024
Accepted December 2nd, 2024
Published December, 2024

***Abstract*— Effective customer segmentation is crucial in optimizing marketing strategies, particularly in customer-oriented aesthetic clinics. This research aims to enhance customer segmentation in aesthetic clinics using a K-Means approach based on the RFAM (Recency, Frequency, Average-Monetary) model. This approach is utilized to leverage historical customer data to identify customer segments based on their purchasing behavior, including visit frequency, average purchase amount, and the last time they visited the clinic. The K-Means clustering method maps customers into homogeneous groups, enabling aesthetic clinics to adapt more focused and personalized marketing strategies. The research results indicate insights obtained from the analysis and interpretation of RFAM conducted on 493 data points, resulting in the formation of two distinct clusters. In Cluster 1, denoting low loyalty, there are 156 customers, while Cluster 2 comprises 337 customers, reflecting high loyalty. Practical implications of this research include improvements in service customization and promotions tailored to customer needs and preferences. In conclusion, the K-Means approach based on the RFAM model can be utilized as an effective tool to enhance customer segmentation in the aesthetic clinic industry.**

***Keywords*—clustering; customer segmentation; historical customer data; marketing strategies; service customization**

1 INTRODUCTION

This research emphasizes the importance of understanding customer loyalty levels as a critical factor in business strategy. Knowing whether high or low loyalty enables businesses to identify key customer segments, tailor personalized marketing strategies, and enhance customer retention and satisfaction. In the healthcare service industry, particularly in aesthetic clinics, while understanding customer preferences and needs is crucial for increasing loyalty and satisfaction, this study specifically focuses on segmenting customers based on their loyalty levels, identifying whether their loyalty is low or high. Customer segmentation is essential for comprehending the variations among consumers, enabling the provision of more focused and personalized services [1].

By identifying different customer groups, clinics can tailor their services to meet specific demands, enhancing customer satisfaction and loyalty [2]. This approach improves service quality and helps clinics stay competitive in a rapidly evolving market. Effective customer segmentation leads to better resource allocation and more targeted marketing efforts, ultimately contributing to the clinic's success and growth [3]. A challenge currently faced by aesthetic clinics lies in comprehensively understanding the preferences and needs of their customers. With the emergence of various beauty treatments and products, coupled with individual aesthetic preferences, aesthetic clinics need to develop more sophisticated customer segmentation strategies to meet optimally customer needs [4].

In this context, the primary issue is the lack of sophisticated and detailed customer segmentation in aesthetic clinics. The problem formulation focuses on improving customer segmentation to make it more accurate and effective [5]. Enhanced segmentation methods can help clinics better understand their client's diverse needs and preferences [6]. This improvement allows for more personalized and targeted services, leading to higher customer satisfaction [7]. By refining these approaches, clinics can allocate resources more efficiently and implement more precise marketing strategies [8]. Ultimately, better customer segmentation can drive the clinic's success and foster stronger customer loyalty [9].

Acknowledging this challenge, efforts are directed towards developing comprehensive segmentation techniques that cater to the unique demands of aesthetic clinic clientele [10]. Collaborative endeavors with industry experts and researchers are essential for advancing segmentation methodologies and addressing the evolving needs of the market [11]. Additionally, leveraging advanced analytics and machine learning algorithms can enhance the precision and scalability of segmentation models [12]. Through continuous refinement and adaptation, aesthetic clinics can achieve a deeper understanding of their customer base, enabling them to deliver unparalleled experiences and stay ahead in a competitive landscape [13].

This study aims to advance our understanding of customer loyalty by providing a detailed segmentation based on loyalty levels, utilizing the K-Means clustering algorithm combined with the RFAM model. It seeks to explore the distinction between high and low-loyalty customers, offering valuable insights into targeted marketing strategies. Hopefully, it would contribute to the body of knowledge by applying an innovative approach to customer segmentation and offering practical implications for businesses aiming to enhance customer retention [14].

This research does not aim to explore customer preferences and needs; instead, it focuses on understanding and segmenting customers based on their levels of loyalty. While preferences and needs may influence loyalty, this research specifically analyzes customer loyalty levels as a key factor for segmentation. This study's results were compared to those of similar customer segmentation studies, demonstrating a higher accuracy in segmenting customer loyalty levels. The K-Means clustering combined with the RFAM model produced more precise loyalty segments, as indicated by the superior Silhouette score (compared to previous studies using simpler clustering methods such as k-means alone) [15]. This enhanced segmentation empowers clinics to deliver more personalized and effective services and to enhance customer satisfaction and loyalty.

Ultimately, the research endeavors to furnish aesthetic clinics with a robust tool for optimizing resource allocation and implementing targeted marketing strategies, thereby fostering overall clinic success [16]. Through rigorous analysis and implementation, it is hoped to elevate the standard of customer segmentation in aesthetic clinics, providing actionable insights for improving service delivery and enhancing customer experiences [17]. The expected benefits may include a profound impact on aesthetic clinics, elevating their customer service strategies to new heights, and a more precise result compared with other customer segmentation studies focused on loyalty. Employing the RFAM model, which offers higher accuracy and more actionable insights, would also be beneficial.

Unlike previous studies, which often rely on general clustering methods, this research combines K-Means with RFAM to deliver more refined loyalty segments, enabling better-targeted marketing strategies. This tailored approach fosters stronger, long-lasting client relationships, and nurtures loyalty over time. Insights gained from distinct customer needs and preferences help clinics optimize resource allocation and implement targeted marketing initiatives effectively. Ultimately, this research aims to empower aesthetic clinics to achieve heightened customer loyalty and enduring success in an increasingly competitive market. The overarching goal is cultivating a responsive, proactive environment that meets customers' demands. Through continuous refinement and adaptation, aesthetic clinics can establish themselves as leaders in customer-centric service delivery. Such advancements are essential for clinics to thrive in today's dynamic and ever-changing business landscape. By embracing the findings of this research, aesthetic clinics can position themselves for sustained growth and prosperity in the years to come.

This research proposes combining customer segmentation approaches using the K-Means method with the RFAM model to obtain more detailed and relevant customer clusters. This study does not include demographic factors or explore customer preferences and needs as variables. It focuses solely on analyzing the level of customer loyalty, aiming to provide a clear segmentation based on loyalty levels without delving into demographic or behavioral determinants, instead [18].

The segmentation in this study specifically categorizes customers based on their loyalty levels.

Customers are divided into distinct segments such as 'low loyalty' and 'high loyalty,' determined by factors like frequency of visits, purchase consistency, and engagement metrics. As a result, this study focuses exclusively on analyzing customer loyalty rather than customer satisfaction. While the introduction highlights the potential impacts of increased loyalty and satisfaction, this research does not evaluate satisfaction as a variable but instead segments customers based on their loyalty levels. This approach ultimately strengthens their competitive position in the market, ensuring they meet diverse customer demands more effectively.

2 METHOD

This research involves several stages to create a sales grouping system determining customer loyalty levels. The first stage is preparing primary data, ensuring accurate and comprehensive data collection. All relevant aspects for subsequent analysis are covered by the data collected. Next, the RFAM analysis stage is undertaken, where data on Recency (when the customer last transacted), Frequency (how often the customer transacts), and Average-Monetary value (the average transaction value of the customer) are analyzed in-depth. Following the RFAM analysis, segmentation is performed using the K-Means Clustering method [4]. Customers are grouped based on their behaviors, allowing customer groups with different characteristics to be identified. The results of this segmentation are then visualized to provide clear insights into the various customer segments.

In the final stage, testing and evaluation are conducted. The testing and evaluation of the model were carried out using a 70-30 train-test split, where 70% of the labeled data was used for training and 30% for testing. the number of samples used for testing was 147, based on the 70-30 split of the 493 total records. The model's performance was evaluated using metrics, such as accuracy, silhouette score, and F1 score to ensure reliable results. The effectiveness and accuracy of the segmentation performed are assessed during this stage. Ensuring that the created grouping system can provide useful results for customer loyalty assessment is crucial.

Overall, these stages aim to enhance customer loyalty assessment and service strategies. By understanding customer groups more deeply, companies can develop more targeted plans to improve customer satisfaction and loyalty. Implementing this system is expected to result in recognizing loyal customers and designing more effective loyalty programs. Consequently, sales are boosted, and long-term relationships with customers are strengthened.

2.1 Dataset

The dataset utilized comprises transactional data spanning six months from June 2023 to December 2023, totaling 493 records. In Table 1 (raw dataset), the following attributes are included: Date, Invoice No., Product, Qty, Total, Acc. No, and Customer Name. Businesses can learn a great deal about the performance of their goods and services, consumer preferences, industry trends, and possible areas for improvement in their marketing strategy by closely examining sales data. Making well-informed company decisions that are focused on the demands of the market is made easier with the use of sales data, which boosts productivity and leads to long-term success [19]. The dataset's details are displayed in Table 1.

2.2 Data Cleaning

The dataset intended for the transaction grouping system is cleansed in this stage. The process begins with the pre-processing stage, where certain attributes are removed from the dataset. For example, attributes such as 'customer age' and 'address' were excluded due to their low correlation with customer loyalty levels, as indicated by preliminary feature selection analysis and correlation tests. The decision to remove these attributes was based on the principle of reducing dimensionality without losing critical information relevant to loyalty segmentation, which allows the relevant information for analysis to be focused upon [20].

Next, errors are corrected. Outlier analysis was conducted using the Z-score method to identify extreme values in the dataset. Any data points with a Z-score above 3 or below -3 were considered outliers. These outliers were removed from the dataset to ensure that the model focused on more representative customer behavior and to prevent skewing of the clustering results, which can distort analysis results if not properly managed. Data entries with significant deviations or inaccuracies are identified and either corrected or removed. Thus, the transaction patterns will accurately be represented by the dataset.

the data cleansing process Standardizing formats and ensuring data consistency are also included in. For example, dates are uniformly formatted, and monetary values are standardized to a common currency or unit. This step is vital to prevent discrepancies that could arise during analysis due to inconsistent data formats. Ensuring uniformity in data entries allows for smoother and more accurate processing in subsequent stages. Additionally, duplicate records are identified and eliminated to avoid redundant information that could skew the analysis results [21].

By meticulously cleaning the data, the accuracy and relevance of the subsequent analysis and segmentation are greatly improved. To support the findings of this research, comparative data from similar studies were analyzed. For instance, the results of this study were compared with those of previous customer loyalty segmentation studies, which used different methods like K-means clustering without the RFAM model.

Table 1. Raw Dataset

Date	**Invoice No.**	**Product**	**Qty**	**Total**	**Acc. No**	**Customer Name**
6/1/2023	IN-00006469	BOTOX A 10 Unit	5	3975000	SB00 018	SXXXXX X
…….	……..	………	……	……….	…… ….	…….
12/31/20 23	IN-00007540	ACNE GEL	1	215000	SB01 6073	AXX SXXX

These comparisons demonstrated the superior accuracy and precision of our approach to segmenting customer loyalty. The effectiveness of the customer segmentation and loyalty assessment system is enhanced by this crucial step. A reliable foundation for the RFAM analysis stage, where Recency, Frequency, and Average-Monetary values are calculated, is provided by clean data. The precision of the customer segmentation process is directly impacted by the accuracy of these calculations. Additionally, the performance of the K-Means Clustering method used for segmentation is enhanced by well-prepared data. The quality of the input data determines the quality of the clusters formed, which in turn affects the insights drawn from the segmentation. A better understanding of customer behaviors and more effective targeting of loyalty programs are enabled by reliable and accurate clusters. Overall, ensuring that the dataset is free from errors and inconsistencies is the aim of the data cleansing stage, providing a solid base for further analysis. Precise segmentation, essential for developing effective customer loyalty strategies, is achieved through this meticulous preparation. The outcomes of customer loyalty assessment initiatives are significantly enhanced by thorough data cleansing, improving customer satisfaction and long-term business success.

2.3 RFAM Model

Before proceeding with the segmentation process using clustering, it is necessary to designate attributes as values for Recency (R), Frequency (F), and Average Monetary (AM) in the dataset [22]. This involves selecting specific data points that accurately represent how recently a customer has made a purchase (Recency), how often they make purchases (Frequency), and their average spending amount (Average Monetary) [23]. By defining these attributes, the dataset is prepared for more precise and meaningful segmentation [24]. This study does not analyze customer behavior; but focuses on segmenting customers based on their loyalty levels. The research aims to identify groups of customers with varying loyalty levels but does not explore personalized targeting strategies or behaviors associated with these segments. Targeted and personalized strategies based on customer loyalty will be considered in future research.

Recency (R) measures the time elapsed since a customer's last transaction, indicating how recently they made a purchase [25]. This metric assumes that customers who have made recent purchases are more likely to respond to marketing efforts. A customer who completes the purchase more recently is generally considered more engaged and valuable than one who has not purchased in a while. One effect of analyzing the recency is that businesses can better understand customer engagement levels and tailor marketing strategies accordingly. This metric is crucial for assessing customer activity and determining the effectiveness of retention efforts. Understanding recency helps businesses identify and prioritize customers who are most likely to generate revenue.

Frequency indicates how often a customer makes purchases within a defined timeframe, reflecting their loyalty and engagement with the brand [26]. Customers who make frequent purchases tend to be more loyal and contribute more to the brand's revenue over time. Segmentation based on purchase frequency allows businesses to effectively tailor their marketing strategies to different customer segments. By identifying high-frequency purchasers, businesses can prioritize efforts to nurture and retain these valuable customers. Understanding purchase frequency helps businesses gauge customer activity levels and adjust strategies accordingly. This segmentation metric is crucial for optimizing marketing efforts and maximizing customer lifetime value.

Average Monetary represents the average amount spent by a customer during a specific period, indicating their overall contribution to revenue and profitability [27]. Customers with higher AMVs are often regarded as high-value customers and may receive special incentives to encourage repeat purchases. Understanding AMV allows businesses to identify and prioritize customers who have the potential to generate more revenue. By segmenting customers based on their AMVs, businesses can tailor marketing strategies to maximize profitability. This metric helps businesses assess the effectiveness of their pricing strategies and promotional efforts. AMV is a crucial factor in customer segmentation, guiding businesses in allocating resources and implementing targeted marketing initiatives.

2.4 Normalization

The predetermined RFAM scores will undergo a normalization stage for each RFAM attribute using Equation 1.

$$v' = \frac{v - min_a}{max_a - min_a}(newmax_a - newmin_a) + newmin_a \quad (1)$$

v' is the normalized value, where v represents the value before normalization, min_a is the minimum value for each variable, max_a is the maximum value for each variable, $newmax_a$ is the maximum range for x with a value of 1, and $newmin_a$ is the minimum range for x with a value of 0.

2.5 Silhouette Coefficient

The Silhouette Method employs a silhouette coefficient [28] to assess both the separation and cohesion within clusters. This coefficient is calculated by dividing the separation measure by the cohesion measure and then subtracting 1 if the separation measure exceeds the cohesion measure. Conversely, if the cohesion surpasses the separation, 1 is subtracted from the cohesion measure divided by the separation measure. A higher Silhouette coefficient indicates a more favorable clustering outcome, reflecting the quality and appropriateness of the clusters generated by the algorithm. Through evaluating both separation and cohesion, the Silhouette Method provides valuable insights into the effectiveness of the clustering process, aiding in the interpretation and validation of results.

This method serves as a critical tool for researchers and analysts seeking to optimize clustering techniques and enhance the accuracy of segmentation outcomes. By leveraging the Silhouette Method, practitioners can refine their clustering algorithms and make more informed decisions based on the quality of the clusters produced. Additionally, the Silhouette coefficient offers a standardized metric for comparing clustering results across different datasets and algorithms, facilitating robust and consistent evaluation. Overall, the Silhouette Method is pivotal in validating clustering techniques, contributing to the advancement of data-driven decision-making and insights generation (Equation 2).

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{2}$$

Where S_i is the Silhouette index, a_i represents the average distance between point i and all other points within cluster A (the cluster where point i_i belongs), and b_i is the average distance between point i and all other points in clusters other than A.

2.6 K-Means Clustering

Clustering is recognized as an unsupervised learning technique pivotal in dividing a heterogeneous population into more homogeneous groups, with applications found across various domains such as customer segmentation, image segmentation, information retrieval, web page clustering, scientific analysis, and analytical techniques [29], [30], [31], [32]. It operates without predefined labels, making it more adaptable to diverse datasets and objectives. Among clustering methods, K-Means is widely utilized in distance-based clustering algorithms, where data is effectively partitioned into a specified number of clusters, predominantly operating on numeric attributes.

In the critical phase of customer segmentation, the optimal value of k, representing the number of clusters to be formed, must be determined [33]. This is often facilitated by using the elbow method, where the within-cluster sum of squares is plotted against the number of clusters. The 'elbow' point on the graph indicates a diminishing return in variance reduction, helping researchers discern the ideal number of clusters from the dataset. This ensures that the segmentation is neither granular nor coarse, providing a balanced view of customer groupings. In K-Means Clustering, the proximity of objects to centroids is calculated, typically through Euclidean Distance. This distance measure involves an equation that computes the straight-line distance between paired data objects in multidimensional space, and by minimizing the total distance within clusters, centroids are iteratively adjusted to better represent the data points.

The rigor and systematic approach adopted in clustering techniques are underscored by the utilization of such mathematical methodologies [34]. Robust and insightful outcomes across diverse applications are ensured by providing a clear mathematical basis for cluster formation and adjustment. In the broader context of unsupervised learning, the meticulous calculation of distances and allocation of data points to appropriate clusters are integral processes. These processes foster a deeper understanding of complex datasets and facilitate informed decision-making by revealing hidden patterns and relationships within the data.

As clustering techniques evolve, indispensable tools in extracting meaningful patterns and structures from vast and disparate datasets are created. Innovations in algorithms and computational power further enhance their accuracy and applicability, driving progress in numerous fields and industries worldwide. The profound and far-reaching impact of clustering, from marketing strategies tailored to specific customer segments to groundbreaking discoveries in scientific research, is recognized. By continuously refining these techniques, new levels of insight and efficiency can be unlocked by researchers and practitioners, propelling innovation and success across various domains.

3 RESULT AND DISCUSSION

3.1 RFAM Transformation

The RFAM transaction dataset serves as a crucial resource for understanding member behaviors and spending habits at the clinic. It provides detailed insights into how frequently members engage with the clinic's services and their financial commitment. This data is essential for identifying trends and optimizing marketing strategies tailored to member needs. The information presented in Table 2 is the RFAM transaction dataset, which includes unique member identification (Member ID), the time elapsed since the member's last transaction (Recency) in days, the number of transactions made by the member (Frequency), and the average amount of money spent per transaction (Average Monetary). By utilizing this table, patterns of member spending, transaction frequency, and average expenditure tendencies at the clinic can be analyzed.

3.2 Normalization

Normalization is a critical process in data pre-processing, ensuring that different attributes contribute equally to analyses. By standardizing the range of numeric values, we can enhance the performance of machine learning algorithms. This technique is particularly useful when dealing with diverse scales among features, as it prevents any single attribute from dominating the model. Additionally, normalization facilitates better convergence in optimization processes. The normalization technique entails converting numeric attributes into a narrower range, usually bounded by 0 as the minimum and 1 as the maximum. Below is the computation for Min-Max normalization, as depicted in Table 3.

Table 2. RFAM Transaction Dataset

No	Member ID	Recency	Frequency	Average Monetary
1	SB-000033	94	1	4560000
2	SB-000060	125	6	1375000
3	SB-000081	199	12	4222991.667
4	SB-000082	201	3	952000
5	SB-000120	38	4	1460000
6	SB-000121	195	2	10334000

7	SB-000125	125	3	830416.6667
8	SB-000150	211	10	1312500
9	SB-000158	134	15	309200
10	SB-000236	14	186	4080464.286
…	…….	…		………
493	SB-016677	1	4	5600000

3.3 *Silhouette Score*

In clustering, selecting the appropriate number of clusters is pivotal for meaningful data analysis. One effective method for this is the silhouette score, which quantifies the quality of clusters formed. We gain insights into clustering effectiveness by assessing how similar an object is to its cluster compared to others. The silhouette score aids in visualizing cluster cohesion and separation, providing a quantitative measure for evaluation. The Silhouette score calculation is performed after the clustering process has been completed, based on the labels assigned to the clusters.

This metric is used to evaluate the cohesion and separation of the clusters, ensuring that the model has effectively grouped similar data points while maintaining distinct separation between different clusters. This score indicates how well the data points are clustered, ranging from -1 to 1. Such evaluation is aided in determining the most suitable number of clusters by selecting the number that yields the highest silhouette score overall. This crucial step ensures that the resulting clusters are aligned with the intrinsic structure of the data, enabling accurate interpretation and usefulness of the cluster results in subsequent analyses. Thus, utilizing the silhouette score as an evaluation metric strengthens the validity of the generated clusters. It then ensures the relevance of clustering to the desired analytical goals.

The Silhouette score measures the effectiveness of clustering, with a range of values from -1 to 1, where higher values indicate better clusters. In the graph, the *X*-axis represents the number of clusters (*k*) ranging from 2 to 10, while the *Y*-axis displays the Silhouette score ranging from 0.40 to 0.55. Data points represent the Silhouette score for each *k* value, with a connecting blue line indicating the trend. The highest score is achieved when *k*=2, reaching 0.5504, signifying the optimality of the two clusters. Following *k*=2, there is a sharp decline until k=5, indicating that increasing the number of clusters does not enhance quality. Although the score increases after *k*=5, it never reaches the same value as when *k*=2, affirming two clusters as the best choice based on this metric. Based on Figure 1, the optimal number of clusters is *k*=2 with a silhouette score of 0.5504. A silhouette value closer to +1 indicates an optimal scenario compared to other clusters.

Table 3. RFAM Normalized Transaction Dataset

No	Member ID	Recency	Frequency	Average Monetary
1	SB-000033	0.436019	0.000000	0.199304
2	SB-000060	0.582938	0.138889	0.053477
3	SB-000081	0.933649	0.305556	0.183874
4	SB-000082	0.943128	0.055556	0.03411
5	SB-000120	0.170616	0.083333	0.057369
6	SB-000121	0.914692	0.027778	0.463669
7	SB-000125	0.582938	0.055556	0.028543
8	SB-000150	0.990521	0.2500000	0.050616
9	SB-000158	0.625592	0.388889	0.004679
10	SB-000236	0.872038	0.361111	0.177348
…	…….	…		………
493	SB-016677	0.009479	0.000000	0.246921

3.4 *K-Means Clustering*

Visualizing data is essential for uncovering patterns and insights that might not be immediately apparent in raw numbers. In the clustering analysis, graphical representations can significantly enhance understanding and communication of results. Figures and charts allow researchers to convey complex relationships among variables in an accessible manner. They also validate the clustering process, demonstrating how well-defined the segments are. Effective visualizations can highlight key differences among groups, making it easier to identify potential strategies for targeted interventions. Additionally, visual tools can facilitate discussions among stakeholders by providing a common reference point for data-driven decisions.

Figures 2, 3 and 4 illustrate the distribution of the two clusters based on Recency, Frequency, and Average Monetary factors. These visualizations portray distinct client clusters along the x-axis and their corresponding RFAM values along the y-axis. The graphical representation enables a clear understanding of the clustering patterns and the relationships between the RFAM variables within each cluster. Cluster 1 and Cluster 2 are discernible based on their positioning and dispersion across the RFAM dimensions, providing insights into the segmentation outcomes.

Outlier analysis was performed before the normalization process. This was done to identify extreme values in the dataset without the influence of scaling, ensuring that outliers were detected in their raw form. After the removal of outliers, the data was then normalized to bring all variables to a comparable scale. By visually examining the distribution of clusters, researchers can glean valuable insights into the characteristics and behaviors of different customer segments. Such visualizations serve as powerful tools for communicating complex data patterns and findings to stakeholders and decision-makers. The graphical depiction of clustering results enhances comprehension and enables more informed decision-making regarding marketing strategies and customer engagement initiatives.

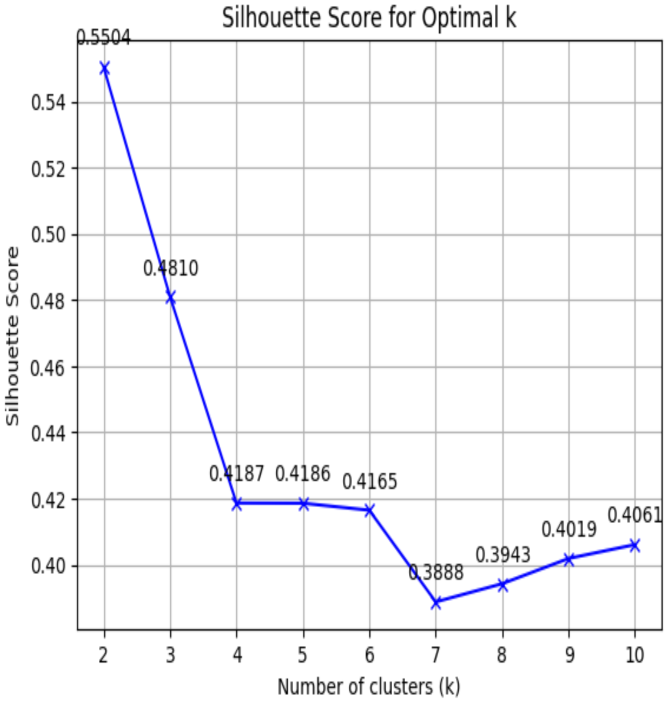

Figure 1. Silhouette score for optimal K

Moreover, visualizations offer a concise and intuitive way to convey the findings of the clustering analysis to a broader audience, including non-technical stakeholders. The use of visual aids in data analysis enhances the interpretability and accessibility of research findings, fostering collaboration and knowledge dissemination. In summary, Figures 2, 3 and 4 provide a visual representation of the clustering results, facilitating a deeper understanding of customer segmentation based on RFAM factors.

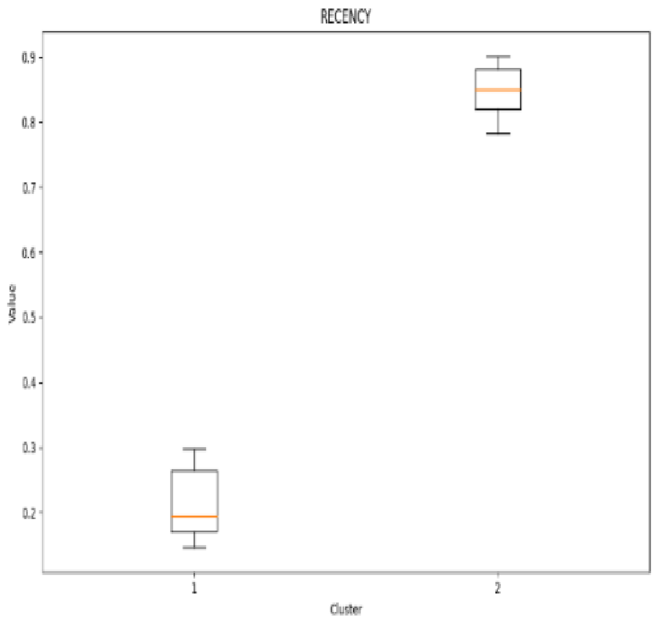

Figure 2. The Cluster's Distribution Based on Recency

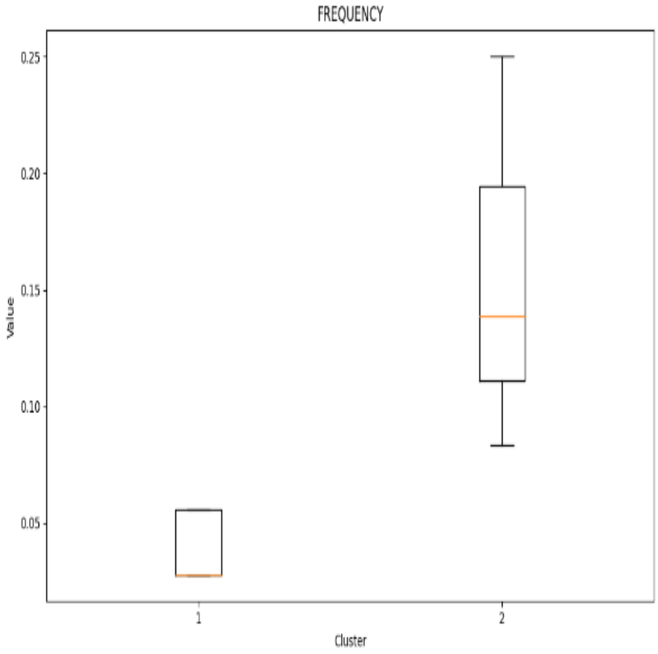

Figure 3. The Cluster's Distribution Based on Frequency

Analyzing cluster sizes is a crucial step in understanding customer behavior. By depicting cluster sizes based on key variables such as Recency, Frequency, and Average Monetary, we can gain valuable insights into customer segmentation.

A deeper understanding of cluster sizes helps identify the unique characteristics of each segment. This knowledge allows companies to tailor marketing strategies and product offerings to better meet the needs of different groups. Additionally, cluster size analysis can assist in optimizing resource management and business planning.

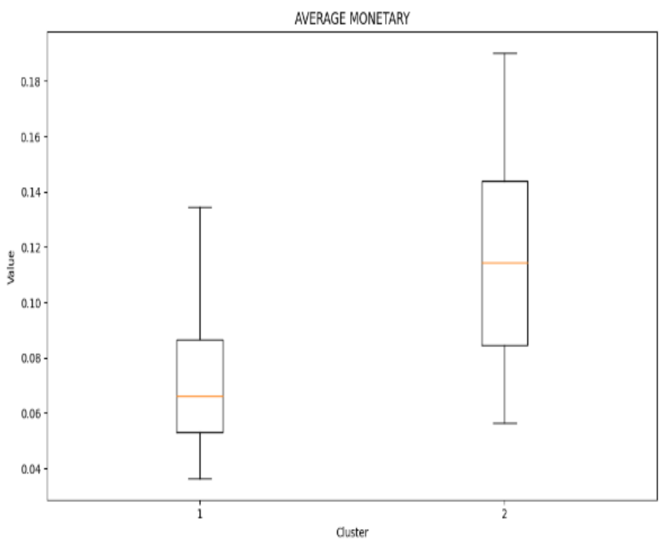

Figure 4. The Cluster's Distribution Based on Average Monetary

Figure 5 illustrates cluster sizes based on Recency, Frequency, and Average Monetary. Cluster Sizes by Recency indicates how often or rarely customers have recently made purchases. If a cluster has a large size in this aspect, it suggests that many customers in that group have made recent transactions, while smaller-sized clusters likely comprise customers who make purchases less frequently. Cluster Sizes by Frequency reflects the frequency of customer transactions within a specific timeframe. Larger clusters in this context may consist of highly active customers who make frequent transactions, whereas smaller clusters may include customers with lower transaction frequencies. Cluster Sizes by Average Monetary denotes the average amount of money spent by customers per transaction. Larger clusters in this category may comprise customers who typically spend more per transaction, while smaller clusters may involve customers who pay less.

Analyzing the relationship between Recency, Frequency, and Average Monetary provides valuable insights into customer behavior. Understanding these dynamics can help identify key segments that contribute significantly to revenue. By recognizing patterns among high-value customers, businesses can develop targeted marketing strategies. Additionally, examining these relationships allows for identifying emerging trends, such as new customers who may become loyal patrons. This strategic approach can optimize customer engagement efforts and drive growth.

Based on Figure 6, customers who exhibit both high Recency and high Frequency tend to demonstrate the highest Average Monetary value. Nevertheless, certain customers with either low Recency or low Frequency display elevated Average Monetary values. This suggests that a subset of new customers is making substantial purchases. This insight can be leveraged to pinpoint customers with significant potential for repeat purchases and enhance their overall purchases value.

Understanding customer behavior through data analysis is vital for any business looking to thrive in a competitive landscape. Effective segmentation allows companies to tailor their offerings and marketing strategies to specific groups. By examining patterns in customer transactions, businesses can identify trends that inform their approach. This process not only enhances customer engagement but also drives overall business performance. Leveraging data insights fosters a culture of informed decision-making, empowering teams to respond proactively to market demands.

Table 4 displays the outcomes of cluster analysis centered on the RFAM (Recency, Frequency, Average Monetary) variable. The initial cluster, comprised of 156 data points, signifies customers with low transaction frequency and recency yet high average transaction value. Conversely, the second cluster, containing 337 data points, represents customers exhibiting high transaction frequency, recency, and average transaction value. This delineation facilitates a nuanced understanding of customer behavior and preferences, enabling companies to tailor their strategies accordingly.

By identifying distinct customer segments based on transaction patterns, companies can devise targeted marketing initiatives and refine product or service offerings to better meet the needs of each segment. Such insights empower companies to allocate resources more effectively, optimizing customer engagement and driving business growth. Moreover, understanding the characteristics of different customer segments aids in enhancing customer satisfaction and loyalty, ultimately contributing to long-term success.

Table 4 serves as a valuable tool for decision-makers, providing actionable insights derived from rigorous data analysis. These insights can inform strategic decisions and guide resource allocation, ensuring that companies remain competitive and responsive to evolving market dynamics. Additionally, Table 4 lays the foundation for further research and analysis, offering a basis for exploring correlations between customer segmentation and business performance metrics. By leveraging the insights gleaned from cluster analysis, companies can unlock new opportunities for innovation and growth, positioning themselves for sustainable success in today's dynamic marketplace.

4 CONCLUSION

Insightful results were obtained from the analysis and interpretation of the RFAM, conducted on 493 data points, which led to the formation of two distinct customer clusters. Cluster 1, indicating low loyalty, includes 156 customers, while Cluster 2, reflecting high loyalty, comprises 337 customers. The clustering outcome was validated with an accuracy score of 0.5504, approaching 1, as determined by the silhouette coefficient, indicating a relatively robust clustering result. These findings have significant implications for future research and practical applications in customer segmentation and loyalty assessment.

For future research, it is recommended to conduct comparative analyses with other segmentation methods, which could further enhance the accuracy and effectiveness of the clustering process. Exploring alternative methodologies may offer deeper insights into customer loyalty dynamics and lead to improved segmentation outcomes. Additionally, the generalizability of the findings could be strengthened by examining larger datasets or investigating different industry contexts, thus broadening the understanding of customer behavior across various scenarios. Longitudinal data could also provide valuable insights by tracking changes in customer loyalty over time.

Integrating qualitative research methods—such as interviews or focus groups—could complement quantitative analysis and yield a more comprehensive understanding of the factors that drive customer loyalty. Moreover, considering the impact of external variables, such as economic conditions or industry trends, would provide actionable insights for businesses to adapt their strategies accordingly.

Table 4. Number of Data Points in Each Cluster

Cluster	The Number of Data Points in Each Cluster	RFAM Analysis
Cluster 1	156	↓R ↓F ↑AM
Cluster 2	337	↑R ↑F ↑AM

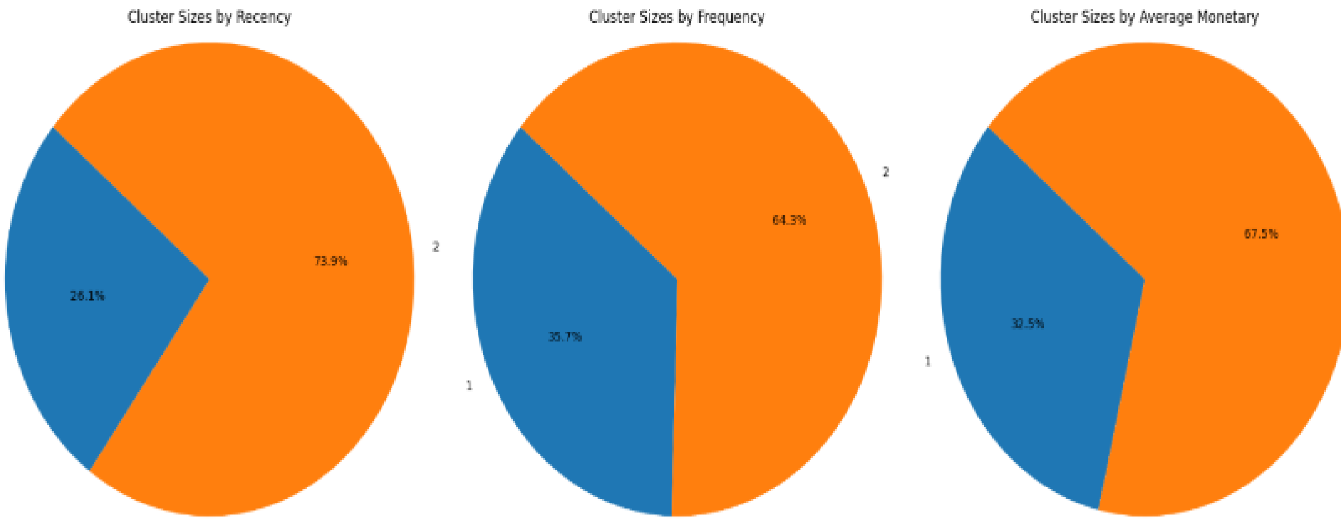

Figure5. Cluster Sizes Based on Recency, Frequency and Average Monetary

The research also benefited from close collaboration with several aesthetic clinics, which provided access to customer records and supported the validation of segmentation models. This partnership enabled the research to address practical challenges within the industry and offer actionable recommendations to enhance customer loyalty management in aesthetic clinics. Exploring cross-cultural differences in customer behavior and preferences could provide valuable insights for businesses operating in diverse markets.

These suggestions aim to deepen the understanding of customer loyalty and improve the practical utility of segmentation techniques, ultimately contributing to enhancing business performance and customer satisfaction. Such efforts would not only advance the field of customer segmentation but also offer valuable strategies for businesses seeking to optimize customer engagement and loyalty.

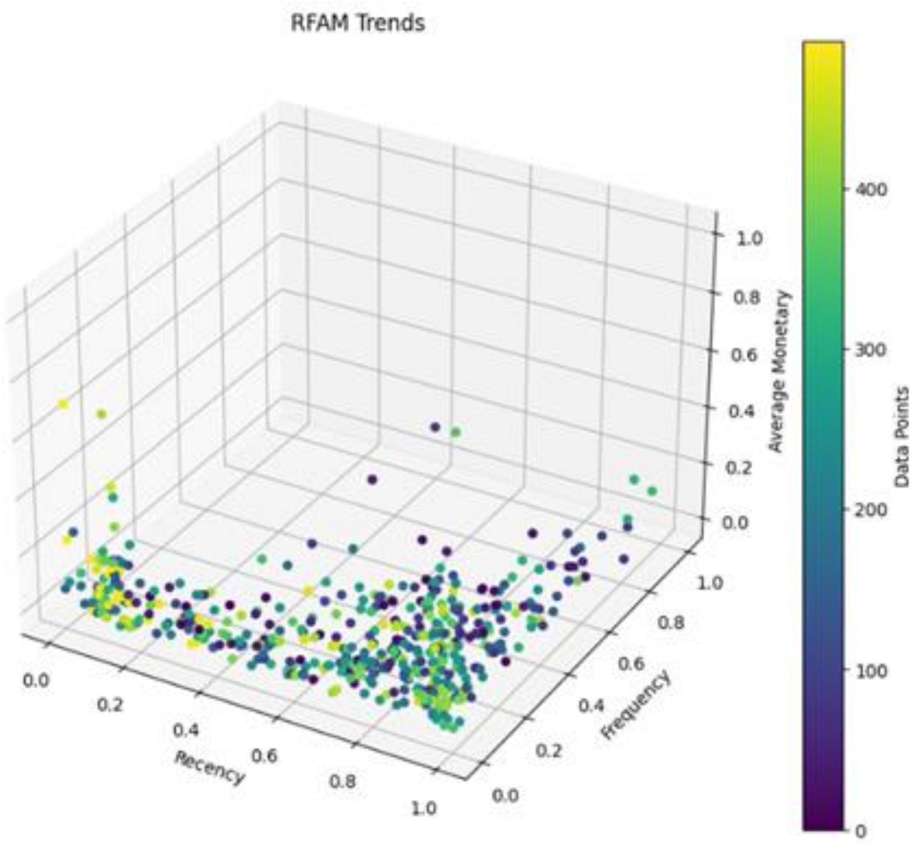

Figure 6. RFAM Trends

AUTHOR'S CONTRIBUTION

The primary author (Sinarring Azi Laga) proposed and designed the research concept, establishing the theoretical framework for utilizing the RFAM model in customer segmentation. The second author (Deny Hermansyah) conducted data analysis and implemented the K-Means algorithm for customer clustering, ensuring an accurate interpretation of segmentation results. The third author (Citra Laksmi Rithmaya) was responsible for data collection from the aesthetic clinic, designing surveys, and relevant data collection instruments. The fourth author (Muhammad Zainuddin) developed and validated the RFAM model used in the study, contributing to the statistical analysis and methodology. The fifth author (Geo Ihsan Ardana) authored the literature review, examining existing research on customer segmentation and pertinent data analysis techniques. Finally, the sixth author (Iqbal Ramadhani) oversaw the writing and editing of the final manuscript, including presenting the research findings at conferences and preparing the paper for publication in academic journals.

COMPETING INTERESTS

According to the publication ethics of this journal, the authors of this article—Sinarring Azi Laga, Deny Hermansya, Chitra Laksmi Rithmaya, Muhammad Zainuddin, Geo Ihsan Ardana, Iqbal Ramadhani M—declare that there are no conflicts of interest (COI) or competing interests (CI) associated with this article.

ACKNOWLEDGMENT

The author sincerely appreciates Hayam Wuruk Perbanas University and the Ministry of Research, Technology, and Higher Education of Indonesia's unwavering support in facilitating and conducting the research. Their collaboration and provision of resources have significantly contributed to the successful completion of this study. The partnership with Hayam Wuruk Perbanas University has provided valuable academic guidance and access to research facilities. Additionally, the Ministry of Research, Technology, and Higher Education of Indonesia's support has been instrumental in navigating regulatory requirements and securing necessary permissions for the research. The author acknowledges the pivotal role played by both institutions in fostering an environment conducive to scholarly inquiry and innovation. Furthermore, their commitment to advancing scientific endeavors has been commendable and deeply appreciated. The author also expresses gratitude to the faculty members and research assistants who contributed their expertise and dedication to the project. Their invaluable insights and diligent efforts have enriched the quality of the research outcomes. Additionally, the author extends thanks to the participants who generously volunteered their time and insights, without whom this study would not have been possible. The author also acknowledges the contributions of colleagues and peers who provided valuable feedback and support throughout the research process. Their encouragement and constructive criticism have been invaluable in shaping the direction of the study. Moreover, the author expresses gratitude to friends and family for their unwavering support and understanding during the research endeavor. Their encouragement and belief in the author's abilities have been a constant source of motivation. Lastly, the author dedicates this research to the pursuit of knowledge and the advancement of science for the betterment of society.

REFERENCES

[1] M. S. Kasem, M. Hamada, and I. Taj-Eddin, 'Customer profiling, segmentation, and sales prediction using AI in direct marketing', *Neural Comput. Appl.*, vol. 36, no. 9, pp. 4995–5005, Mar. 2024, doi: 10.1007/s00521-023-09339-6.

[2] P. S. Paays, 'The Effect of Service Recovery aAccessibility on Customer Satisfaction and Loyalty', *J. Ekon. Bisnis Manaj. Dan Akunt. JEBMA*, vol. 4, no. 1, pp. 539–554, 2024.

[3] A. Dewi Rana, Q. Chloe Milano Hadisantoso, and A. Suganda Girsang, 'RFM-T Model Clustering Analysis in Improving Customer Segmentation', *Int. J. Comput. Digit. Syst.*, vol. 16, no. 1, pp. 1–11, 2024.

[4] S. A. Laga, I. R. Mukhlis, D. Hermansyah, G. Suprianto, M. A. Karyawan, and H. Yutanto, 'Customer Behavior Using RFM Model and K-Means Algorithm in Aesthetic Clinic', in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, IEEE, 2023, pp. 1–5. Accessed: Apr. 23, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10382095/

[5] J. Li, 'How Do Information Quality, E-service Quality, And System Quality Enhance Customer Satisfaction for Airbnb?', *Int. J. Educ. Humanit.*, vol. 13, no. 2, pp. 29–45, 2024.

[6] J. Kim, J. Kim, and S. Hong, 'G-TRACE: Grouped temporal recalibration for video object segmentation', *Image Vis. Comput.*, p. 105050, 2024.

[7] P. Kumar and G. Bisaria, 'Consumer Satisfaction Judgments Based on Credence Services', in *Interdisciplinary Research in Technology and Management*, CRC Press, pp. 481–490. Accessed: May 15, 2024. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003430469-56/consumer-satisfaction-judgments-based-credence-services-pradeep-kumar-gaurav-bisaria

[8] T. K. Vashishth, K. K. Sharma, B. Kumar, S. Chaudhary, and R. Panwar, 'Enhancing Customer Experience through AI-Enabled Content Personalization in E-Commerce Marketing', *Adv. Digit. Mark. Era Artif. Intell.*, pp. 7–32, 2025.

[9] A. Suradi, S. Purwati, M. Zakaria, R. Nurbakti, and F. Mustafa, 'Analysis of Determinant Factors Customer Loyalty Towards Brand in The Telecommunication Industry With The Digitalization Paradigm', *J. Sistim Inf. Dan Teknol.*, pp. 36–41, 2024.

[10] D. Li, Y. Li, Y. Fan, and S. Tang, 'Accurate Segmentation Method for Roadside Lampposts Based on Vehicle-Mounted Lidar Point Clouds', *Int. Arch.*

Photogramm. Remote Sens. Spat. Inf. Sci., vol. 48, pp. 305–310, 2024.
[11] D. S. Ametefe *et al.*, 'Automatic classification and segmentation of blast cells using deep transfer learning and active contours', *Int. J. Lab. Hematol.*, p. ijlh.14305, May 2024, doi: 10.1111/ijlh.14305.
[12] A. Amzil, M. Abid, M. Hanini, A. Zaaloul, and S. El Kafhali, 'Stochastic analysis of fog computing and machine learning for scalable low-latency healthcare monitoring', *Clust. Comput.*, Feb. 2024, doi: 10.1007/s10586-024-04285-x.
[13] Y. Yuricha, B. S. Salim, S. A. Laga, A. Suroni, and R. Permatasari, *Buku Ajar Interaksi Manusia Dan Komputer*. PT. Sonpedia Publishing Indonesia, 2024. Accessed: Oct. 30, 2024. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=HFYrEQAAQBAJ&oi=fnd&pg=PA17&dq=info:TDDnds64A4oJ:scholar.google.com&ots=D5urhI9a9G&sig=9pRxJVWz_RxhqDOHs0bsIPcDLd0
[14] A. Apichottanakul, M. Goto, K. Piewthongngam, and S. Pathumnakul, 'Customer behaviour analysis based on buying-data sparsity for multi-category products in pork industry: A hybrid approach', *Cogent Eng.*, vol. 8, no. 1, p. 1865598, Jan. 2021, doi: 10.1080/23311916.2020.1865598.
[15] F. Yoseph, N. H. Ahamed Hassain Malim, M. Heikkilä, A. Brezulianu, O. Geman, and N. A. Paskhal Rostam, 'The impact of big data market segmentation using data mining and clustering techniques', *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6159–6173, 2020.
[16] D. Hermansyah and A. Muklason, 'Evaluation of hyper-heuristic method using random-hill climbing algorithm in the examination timetabling problem', in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 022101. Accessed: Apr. 23, 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1569/2/022101/meta
[17] M. Furqon, N. Najwa, D. Hermansyah, and M. Zarkasi, 'Knowledge Graph Modeling in Healthcare: A Bibliometric Analysis', *J. Komput. Terap.*, vol. 8, no. 1, pp. 113–122, 2022.
[18] V. V. Muthuswamy and K. Manoharan, 'SDG: Consumer's Perception Towards Herbal/Organic Products with Respect to Sustainable Development Goals', *J. Lifestyle SDGs Rev.*, vol. 4, no. 4, pp. e01994–e01994, 2024.
[19] S. A. Laga, E. T. Sihotang, D. Hermansyah, and T. A. Hariyanti, 'Decision Support System for Selection of An Excellent-Students at Faculty Level in University of X Based on GAP Methods', *J. Adv. Inf. Ind. Technol.*, vol. 6, no. 1, pp. 1–10, 2024.
[20] O. H. Y. Lam, J. Kattge, S. Tautenhahn, G. Boenisch, K. R. Kovach, and P. A. Townsend, '"rtry": An R package to support plant trait data preprocessing', *Ecol. Evol.*, vol. 14, no. 5, p. e11292, May 2024, doi: 10.1002/ece3.11292.
[21] N. M. A. Krishnan, H. Kodamana, and R. Bhattoo, 'Data Visualization and Preprocessing', in *Machine Learning for Materials Discovery*, in Machine Intelligence for Materials Science. , Cham: Springer International Publishing, 2024, pp. 25–46. doi: 10.1007/978-3-031-44622-1_2.
[22] S. Arefin *et al.*, 'Retail Industry Analytics: Unraveling Consumer Behavior through RFM Segmentation and Machine Learning', in *2024 IEEE International Conference on Electro Information Technology (eIT)*, IEEE, 2024, pp. 545–551. Accessed: Oct. 30, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10609927/
[23] S. Kim, W. Shin, and H.-W. Kim, 'Predicting online customer purchase: The integration of customer characteristics and browsing patterns', *Decis. Support Syst.*, vol. 177, p. 114105, 2024.
[24] X. Han, C. Liu, Y. Zhou, K. Tan, Z. Dong, and B. Yang, 'WHU-Urban3D: An urban scene LiDAR point cloud dataset for semantic instance segmentation', *ISPRS J. Photogramm. Remote Sens.*, vol. 209, pp. 500–513, 2024.
[25] M. Țichindelean, C. Ogrean, and M. Herciu, 'DO LOYAL CUSTOMERS BUY DIFFERENTLY? EXAMINING CUSTOMERS'LOYALTY IN A SELF-SERVICE SETTING', *Stud. Bus. Econ. No*, vol. 19, p. 1, 2024.
[26] A. K. Pandey, A. Goyal, and N. Sikka, 'RE-RFME: Real-Estate RFME Model for customer segmentation', Apr. 26, 2024, ***arXiv***: arXiv:2404.17177. Accessed: May 15, 2024. [Online]. Available: http://arxiv.org/abs/2404.17177
[27] S. Monalisa, Y. Juniarti, E. Saputra, F. Muttakin, and T. K. Ahsyar, 'Customer segmentation with RFM models and demographic variable using DBSCAN algorithm', *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 21, no. 4, pp. 742–749, 2023.
[28] S. Paembonan and H. Abduh, 'Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat', *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, p. 48, Sep. 2021, doi: 10.51557/pt_jiit.v6i2.659.
[29] D. Tirthyani, S. Kumar, and S. Vats, 'Clustering and Unsupervised Learning', in *Recent Trends and Future Direction for Data Analytics*, IGI Global, 2024, pp. 119–139. Accessed: Oct. 30, 2024. [Online]. Available: https://www.igi-global.com/chapter/clustering-and-unsupervised-learning/347268
[30] Qomariyah and M. U. Siregar, 'Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering', *JISKA J. Inform. Sunan Kalijaga*, vol. 7, no. 2, pp. 91–99, May 2022, doi: 10.14421/jiska.2022.7.2.91-99.
[31] N. Nugroho and F. D. Adhinata, 'Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa', *Teknika*, vol. 11, no. 3, pp. 170–179, Oct. 2022, doi: 10.34148/teknika.v11i3.502.
[32] M. A. Haq, W. Purnomo, and N. Y. Setiawan, 'Analisis Clustering Topik Survey menggunakan Algoritme K-Means (Studi Kasus: Kudata)'.
[33] C. Rungruang, P. Riyapan, A. Intarasit, K. Chuarkham, and J. Muangprathub, 'RFM model customer

segmentation based on hierarchical approach using FCA', *Expert Syst. Appl.*, vol. 237, p. 121449, 2024.

[34] R. Raman, D. Pattnaik, L. Hughes, and P. Nedungadi, 'Unveiling the dynamics of AI applications: A review of reviews using scientometrics and BERTopic modeling', *J. Innov. Knowl.*, vol. 9, no. 3, p. 100517, 2024.