Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14000

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Retail Marketing Strategy Optimization: Customer Segmentation with Artificial Intelligence Integration and K-Means Clustering

Yuliarni Putri[1)*], Dasril Aldo[2)], Wanda Ilham[3)]
[1)]Universitas Tamansiswa Padang, Indonesia, [2)]Institut Teknologi Telkom Purwokerto, Indonesia, [3)]Universitas Catur Insan Cendekia, Indonesia
[1)] yuliarniputri90@gmail.com, [2)]dasrilaldo@telkomuniversity.ac.id, [3)]wandailham@cic.ac.id,

Submitted :Aug 8, 2024 | **Accepted** : Sept 10, 2024 | **Published** : Oct 2, 2024

Abstract: This study aims to optimize retail marketing strategies through customer segmentation using the K-Means clustering method and RFM (Recency, Frequency, Monetary) analysis. By utilizing transaction data from a large retail company, customers are categorized into six segments: VIP Customers, Loyal Customers, Potential Loyalists, New Customers, At-Risk Customers, and Dormant Customers. This segmentation allows for the implementation of more targeted marketing strategies for each customer group. For example, VIP Customers who represent 3.0% of total customers are very active with significant spending, so they deserve exclusive offers and premium services. Loyal Customers, which account for 7.0% of total customers, show high transaction frequency and loyalty, suitable for loyalty programs and recurring discounts. Potential Loyalists, which comprise 15.0%, show the potential for increased loyalty through retention campaigns. New customers representing 16.3% need a brand recognition and promotion strategy to increase their initial engagement. At-Risk Customers covering 30.7% indicated a decrease in transaction activity and required intervention to prevent churn, while Dormant Customers covering 28.1% required a strong reactivation strategy. The clustering evaluation showed an average Silhouette score of 0.3115, which indicates that the clusters that are formed are quite well defined, although there is still room for improvement. This research provides valuable insights to develop more effective and efficient marketing strategies, as well as increase customer satisfaction and loyalty.

Keywords: Customer Segmentation, K-Means Clustering, RFM Analytics, Retail Marketing Optimization, Artificial Intelligence.

INTRODUCTION

In the ever-evolving digital era, the retail industry faces increasingly complex challenges in understanding and meeting customer needs (Kristanto et al., 2022; Hamdani et al., 2023). With the increasing amount of transaction data available, retail companies have the opportunity to dig deeper into their customer behavior (Akhmetbek, 2022). One effective approach to utilize this data is to use artificial intelligence technology (AI) and clustering techniques for customer segmentation (John et al., 2023; Turkmen, 2022).

Customer segmentation is the process of grouping customers into homogeneous groups based on certain characteristics, such as purchasing behavior, product preferences, and demographics (Chellaboina et al., 2022). With more accurate segmentation, retail companies can develop more personalized and relevant marketing strategies, which in turn can increase customer loyalty and sales (Abidar et al., 2023; Zhang, 2023).

Artificial intelligence offers the ability to analyze data on a large and complex scale with high efficiency (Z. Li et al., 2022). One of the most popular and frequently used clustering algorithms in customer segmentation is K-Means. This algorithm is known for its simplicity and ability to handle large amounts of data (M. Li et al., 2023; Zaib & Ourabah, 2023). K-Means works by dividing data into a number of predetermined clusters based on the similarity between these data (Nie et al., 2023).

Previous research has shown the effectiveness of K-Means in a variety of customer segmentation applications. For example, the application of K-Means for customer segmentation in the e-commerce industry and successfully improved the accuracy of product recommendations (Tabianan et al., 2022). Likewise, the application of K-Means research in shopping behavior analysis shows that usage can identify the most valuable customer segments for

*name of corresponding author

retail companies (Xin & Zhang, 2022). In addition, other research shows that K-Means can be integrated with other machine learning techniques to improve customer churn prediction (Xiahou & Harada, 2022). Subsequent research found that customer segmentation with K-Means helps retail companies develop more targeted marketing strategies, such as tailored promotions and personalized product offers (Wani et al., 2023). Finally, it was revealed that the use of K-Means in retail data analysis can improve operational efficiency and reduce marketing costs (Asuah & Prikutse, 2023).

This research aims to optimize retail marketing strategies through the integration of artificial intelligence and K-Means clustering techniques. Using a dataset of customer transactions from a large retail company, the study will identify patterns and groups of customers based on their shopping behavior. The results of this segmentation will be used to develop more effective marketing strategies, including tailored product recommendations, targeted promotions, and improved customer experience.

The advantage of this study compared to previous studies is the use of larger and more diverse datasets, which cover longer time periods and different types of products. In addition, the study not only focuses on customer segmentation but also integrates the results of segmentation into a more comprehensive and measurable marketing strategy. The methodology used in this study also includes stricter validation measures to ensure the accuracy and relevance of the segmentation results. It is expected that the implementation of optimized marketing strategies based on customer segmentation results can provide a competitive advantage for retail companies in an increasingly competitive market. In addition, the research also aims to contribute to the development of methodologies in retail data analysis and the practical application of artificial intelligence in the industry.

LITERATURE REVIEW

Segmenting customers based on their purchasing behavior is an important task in retail marketing, as it allows companies to tailor their strategies to different customer groups. Over the years, various methods have been proposed and implemented to achieve effective customer segmentation. Among these methods, K-Means clustering has emerged as one of the most popular due to its simplicity and effectiveness in handling large datasets.

A. K-Means Clustering in the E-commerce

Velu and Ravi (2022) applied K-Means for customer segmentation in the e-commerce industry and successfully improved the accuracy of product recommendations. In their research, K-Means was used to segment customers based on purchase frequency, total spending, and the type of product purchased. The results show that the resulting segmentation can help companies develop more targeted marketing strategies and increase customer satisfaction (Tabianan et al., 2022). Another study by Koul and Philip (2021) also showed the successful implementation of K-Means in e-commerce, where they used this algorithm to identify groups of customers with similar shopping behaviors, allowing for more effective personalization (Gautam & Kumar, 2022).

B. Segmentation of Shopping Behavior with K-Means

Christy et al. (2021) showed that the use of K-Means in shopping behavior analysis can identify the most valuable customer segments for retail companies. This study highlights the importance of selecting the right number of clusters to achieve accurate segmentation. Using the Elbow and Silhouette Score methods, this study succeeded in determining the optimal number of clusters to group customers based on their shopping behavior (Christy et al., 2021). Research by Maddumala et al. (2022) also supports these findings, where K-Means is used to segment supermarket customers based on shopping patterns, which helps in the development of more personalized marketing strategies. Furthermore (Maddumala et al., 2022), research by Shirole et al. (2021) confirmed the effectiveness of K-Means in segmenting shopping behavior in the retail environment, showing that this algorithm can identify patterns that are not seen with other methods (Somasekhar et al., 2021).

C. Integration of K-Means with Machine Learning for Churn Prediction

Joung and Kim (2023) showed that K-Means can be integrated with other machine learning techniques to improve customer churn prediction. In this study, K-Means was used to segment customers based on their purchasing patterns, and the results were then used as inputs for the churn prediction model. The results show that this approach can improve the accuracy of churn prediction and assist companies in taking proactive actions to retain customers (Joung & Kim, 2023). Wang (2022) also found that the combination of K-Means and machine learning models can improve the efficiency of churn prediction in the telecommunications sector. Additionally, research by Sabuncu et al. (2020) shows that this integration can reduce churn rates in insurance companies, increasing customer retention.

D. Targeted Marketing Strategy with K-Means

Koul and Philip (2021) found that customer segmentation with K-Means helps retail companies develop more targeted marketing strategies, such as customized promotions and personalized product offers. This research highlights the importance of understanding the characteristics of each customer segment to develop an effective marketing strategy. The segmentation results are used to create detailed customer profiles, which are then used to tailor marketing campaigns (Koul & Philip, 2021). Research by Maddumala et al. (2022) also shows that the use of K-Means in marketing strategies can increase sales conversions and customer

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14000
e-ISSN : 2541-2019
p-ISSN : 2541-044X

loyalty (Maddumala et al., 2022). A study by Mensouri et al. (2022) supports these findings, where they found that segmentation with K-Means increases the effectiveness of digital marketing campaigns (Mensouri et al., 2022).

E. Improving Retail Operational Efficiency through K-Means
Husein et al. (2021) revealed that the use of K-Means in retail data analysis can improve operational efficiency and reduce marketing costs. In this study, K-Means was used to identify the most profitable customer segments and direct marketing resources to those segments. The results show that this approach can increase the return on investment (ROI) of marketing campaigns and optimize the use of company resources (Husein et al., 2021). Dewabharata (2022) found that the application of K-Means in retail inventory management can reduce excess stock and shortage of goods (Dewabharata, 2022).

This study has several advantages that distinguish it from previous research. First, this study uses up-to-date and relevant customer data, so that segmentation results are more accurate and can be directly applied in marketing strategies. This is in line with the findings of Velu and Ravi (2022) which show that the use of actual data can improve the accuracy of product recommendations and customer satisfaction. Second, this study integrates the K-Means method with other machine learning techniques to improve customer churn prediction. This approach has been shown to improve prediction accuracy and assist companies in taking proactive actions to retain customers, as demonstrated by Joung and Kim (2023). Third, the study takes a holistic approach in customer segmentation, not only focusing on the frequency of purchases and total spending, but also considering various aspects such as shopping behavior and the type of product purchased. This approach allows for the development of more personalized and targeted marketing strategies, as identified by Koul and Philip (2021) and Christy et al. (2021). In addition, the study also highlights the importance of operational efficiency by using K-Means to identify the most profitable customer segments, which can further direct marketing resources more effectively, as revealed by Quelal et al. (2021).

METHOD

This study uses the K-Means clustering method for customer segmentation, which is equipped with RFM (Recency, Frequency, Monetary) analysis to identify customer value based on their shopping behavior. This process consists of several main stages shown in the following figure:

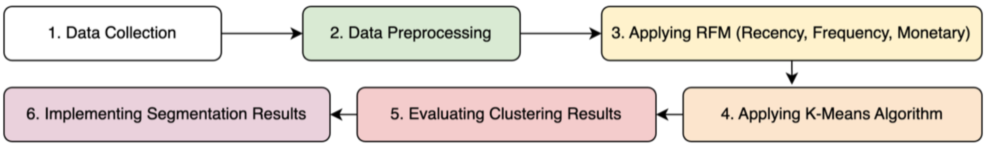

Figure 1. Research Stages

This study uses the K-Means clustering method for customer segmentation, which is equipped with RFM (Recency, Frequency, Monetary) analysis to identify customer value based on their shopping behavior. This process consists of several main stages. The first stage is the collection of customer transaction data from the company's information system, including order_id, customer_id, order_date, product_code, quantity, product_name, and price. The second stage is data preprocessing to eliminate missing or incomplete data, normalize data, and prepare the necessary features. RFM analysis is carried out in the third stage to calculate three main metrics: recency (time since the last transaction), frequency (number of transactions), and monetary (total spending). This metric is used to assess customer value and early segmentation. Furthermore, the K-Means algorithm is applied to group customers into multiple clusters based on their RFM metrics. The clustering results are evaluated using methods such as Silhouette Score to determine the quality of clustering and ensure that customers in one cluster have similar characteristics, while customers in different clusters have different characteristics. The final stage is the implementation of segmentation results in the company's marketing strategy, where the segmentation results are used to develop detailed customer profiles and personalized marketing strategies, such as customized promotions and specific product offers. This approach is designed to ensure that the data used is accurate and relevant, and that the results of segmentation can provide useful insights for better business decision-making. This research is expected to make a significant contribution to the development of more effective and efficient retail marketing strategies.

RESULT

In this study, based on the data analysis that has been carried out, the following is information about the dataset used: The dataset has 499,726 entries and 7 columns. The columns are: order_id, customer_id, order_date, product_code, quantity, product_name, and price. All columns have the corresponding data type (no null values). The dataset is quite large, which is good for customer segmentation analysis. There are 44,693 unique customers,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14000
e-ISSN : 2541-2019
p-ISSN : 2541-044X

indicating a wide customer base. There are 5,688 unique product codes and 5,582 unique product names, indicating the diversity of products sold. The price of the product varies from Rp 800 to more than Rp 1,000,000 (based on statistical values), indicating the diversity in the type and value of the product. Purchase quantities range from 1 to 19 items per transaction.

	order_id	customer_id	order_date	product_code	quantity	product_name	price
0	873411	79795	01/01/2022 08:02	P-8202	5	UNIK SPONGE MOTOR RENCENG	14000
1	873411	79795	01/01/2022 08:02	P-2866	17	DAHLIA DOUBLE BALLK-272	491300
2	873411	79795	01/01/2022 08:02	P-8194	16	DELI BUSA HITUNG UANG / SPONGE TRAY NO 9102	136000
3	685173	34030	01/01/2022 08:02	P-8346	2	RICHEESE SIIP JAGUNG BAKAR 50G	9200
4	685173	34030	01/01/2022 08:02	P-5180	2	LOGEN LABER HARGA NO.111	7000

Figure 2. Raw Data Sample

	order_id	customer_id	order_date	product_code	quantity	product_name	price
0	873411	79795	2022-01-01 08:02:00	P-8202	0.2222222222	UNIK SPONGE MOTOR RENCENG	0.0008370748
1	873411	79795	2022-01-01 08:02:00	P-2866	0.8888888889	DAHLIA DOUBLE BALLK-272	0.0311049387
2	873411	79795	2022-01-01 08:02:00	P-8194	0.8333333333	DELI BUSA HITUNG UANG / SPONGE TRAY NO 9102	0.0085736753
3	685173	34030	2022-01-01 08:02:00	P-8346	0.0555555556	RICHEESE SIIP JAGUNG BAKAR 50G	0.000532684
4	685173	34030	2022-01-01 08:02:00	P-5180	0.0555555556	LOGEN LABER HARGA NO.111	0.0003931715

Figure 3. Data After Normalization

In Figure 2 is the raw data used, then the normalization process is carried out in the 'quantity' and 'price' columns using MinMaxScaler, the results are seen in Figure 3. It converts the values in those columns into a range of 0 to 1, which helps in further analysis. Furthermore, an analysis will be carried out on the Implementation of RFM (Recency, Frequency, Monetary) Calculating recency as the number of days since the last transaction. Calculate frequency as the total number of transactions made by customers. Calculate monetary as total customer spending.

	recency	frequency	monetary
customer_id			
3	103	18	0.167706668695938
4	383	16	0.1160046165943729
5	101	16	0.133316845496283
8	455	9	0.131363670953504
11	381	10	0.0989270222966283

Figure 4. Results of RFM Analysis

Figure 4 shows the results of RFM (Recency, Frequency, Monetary) distribution using the Min-Max Scaling method to provide in-depth insights into customer behavior based on their transaction data. The recency distribution shows that most customers made their last transaction in the last 100 days, with the number of customers gradually decreasing as time has increased since the last transaction. This shows that many customers are still actively transacting, although there are also customers who have not made transactions for a long time. The frequency distribution shows that most customers have a low transaction frequency, which is less than 10 transactions. Only a few customers have a high transaction frequency, which indicates a general distribution pattern where many customers make few transactions and only a few customers transact frequently. The monetary distribution shows that most customers have relatively low total spending, with the distribution decreasing as total spending increases. Just like frequency distribution, only a few customers have a very high total spend, which means that most of the revenue comes from a small percentage of customers.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14000
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	recency	frequency	monetary_normalized
3	103	18	0.1677066687
4	383	16	0.1160046166
5	101	16	0.1333168455
8	455	9	0.131363671
11	381	10	0.0989270223

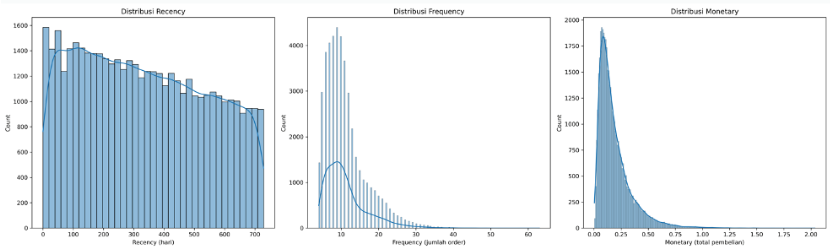

Figure 5. RFM Results

Figure 5 presents the results of RFM (Recency, Frequency, Monetary) analysis in the form of tables and data distributions that will be used as the basis for customer segmentation using the K-Means algorithm. The RFM table displays several important metrics: recency shows the number of days since the customer's last transaction, frequency shows the total number of transactions made by the customer, and monetary_normalized shows the total normalized spend. For example, a customer with ID 3 has a recency of 103 days, a frequency of 18 transactions, and a monetary_normalized of 0.1677. The recency distribution shows that most customers made their last transaction in the last 100 to 200 days. The frequency distribution shows that most customers make less than 10 transactions, and the monetary distribution shows that most customers have a relatively low total spend.

After the RFM data is analyzed, the K-Means algorithm is applied to customer segmentation. This process begins with normalizing the data to ensure that each metric has an equal weight. The Elbow method is used to determine the optimal number of clusters, which in this case is 6 clusters. The K-Means algorithm then groups customers based on similarities in their RFM metrics, with an iterative process to update the centroid position and regroup the customer until the centroid position is stable. Each cluster is analyzed to understand its characteristics and named according to customer profiles, such as VIP Customers, Loyal Customers, Potential Loyalists, New Customers, At-Risk Customers, and Dormant Customers. With this process, more precise and useful customer segmentation can be obtained, allowing companies to develop more effective and targeted marketing strategies.

	recency	frequency	monetary
0	552.7805752529684	7.8302127320531465	0.09862513043780541
1	194.39342147694313	25.100612705578875	0.4206248576746647
2	187.5081871770611	7.971617786187199	0.1041894083267496
3	150.96040148494518	14.726522755396593	0.2452844490893605
4	288.9558052434457	16.146816479400776	0.7286894698060715
5	500.65361625821834	12.763299462044259	0.2595741105039715

Figure 6. Centroid Point Result for K=6

Figure 6. shows the centroid value for each cluster based on RFM (Recency, Frequency, Monetary) analysis, which can be used to segment customers into different segments. Based on this analysis, Cluster 0 is named "Dormant Customers" with a recency value of 552.78 days, transaction frequency of 7.83, and monetary 0.0986, indicating that customers in this cluster tend to have not made transactions for a long time and have low spending value. Cluster 1 is named "VIP Customers" with a recency value of 194.39 days, transaction frequency of 25.10, and monetary 0.4206, indicating that these customers are very active, frequent transactions, and have a high spending value. Cluster 2 is called "New Customers" with a recency value of 187.50 days, a transaction frequency of 7.97, and a monetary value of 0.1042, describing customers who are quite active but with a lower spending value. Cluster 3 is "Loyal Customers" with a recency value of 150.96 days, transaction frequency of 14.73, and monetary 0.2453, which shows that these customers are quite valuable and active. Cluster 4 is called "Potential Loyalists" with a recency value of 288.96 days, a frequency of 16.15 transactions, and a monetary of 0.7287, which shows that they are quite active and highly valuable customers. Finally, Cluster 5 is named "At-Risk Customers" with a recency value of 500.65 days, transaction frequency of 12.76, and monetary 0.2596, which indicates that

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14000
e-ISSN : 2541-2019
p-ISSN : 2541-044X

these customers need special attention to be reactivated. By using these segment names, companies can segment their customers more effectively and develop marketing strategies that better suit the characteristics and values of each cluster. Segments such as VIP Customers and Loyal Customers can receive exclusive offers and loyalty programs, while Dormant Customers and At-Risk Customers can be reactivated with more aggressive marketing campaigns.

Cluster	recency	frequency	monetary	Label
0	552.9007092198582	7.831380986532792	0.09865844038642915	Dormant Customers
1	194.28273427471117	25.066431322207958	0.4198820817433638	Loyal Customers
2	187.64268026219958	7.967662053896577	0.10408320976000102	At-Risk Customers
3	150.85258525852586	14.703932893289329	0.24493970738121476	New Customers
4	288.75973053892216	16.17814371257485	0.7290766676923142	VIP Customers
5	500.5269058295964	12.766218236173392	0.2596359877822031	Potential Loyalists

Figure 7. Results of Cluster Member Grouping

Figure 7 clustering shows the grouping of customers into six segments based on RFM (Recency, Frequency, Monetary) analysis, which allows companies to develop more effective and targeted marketing strategies. The first segment, "VIP Customers," consisted of 1336 customers who were quite active with a recency value of 288.76 days, a frequency of 16.18 transactions, and a monetary of 0.7291. These customers often transact and have a very high value of expenses, making them very valuable to the company. "Loyal Customers," with 3,116 customers, had a recency of 194.28 days, a transaction frequency of 25.07, and a monetary of 0.4199, indicating that they are very loyal and shop frequently. "Potential Loyalists," with 6690 customers, have a recency of 500.53 days, a transaction frequency of 12.77, and a monetary of 0.2596, indicating the potential to become more loyal with the right strategy. "New Customers," which totals 7,272 customers, has a recency of 150.85 days, a transaction frequency of 14.70, and a monetary of 0.2444, indicating that they are just starting to transact and need an approach to increase their engagement. "At-Risk Customers," with 13,730 customers, had a recency of 187.64 days, a transaction frequency of 7.97, and a monetary of 0.1048, showing signs of declining activity and requiring special attention. Lastly, "Dormant Customers," which totals 12549 customers, has a recency of 552.90 days, a transaction frequency of 7.83, and a monetary of 0.0987, indicating that they have not transacted for a long time and need a reactivation strategy. With this segmentation, companies can allocate their marketing resources more efficiently and effectively, targeting customers according to the characteristics and values of each segment.

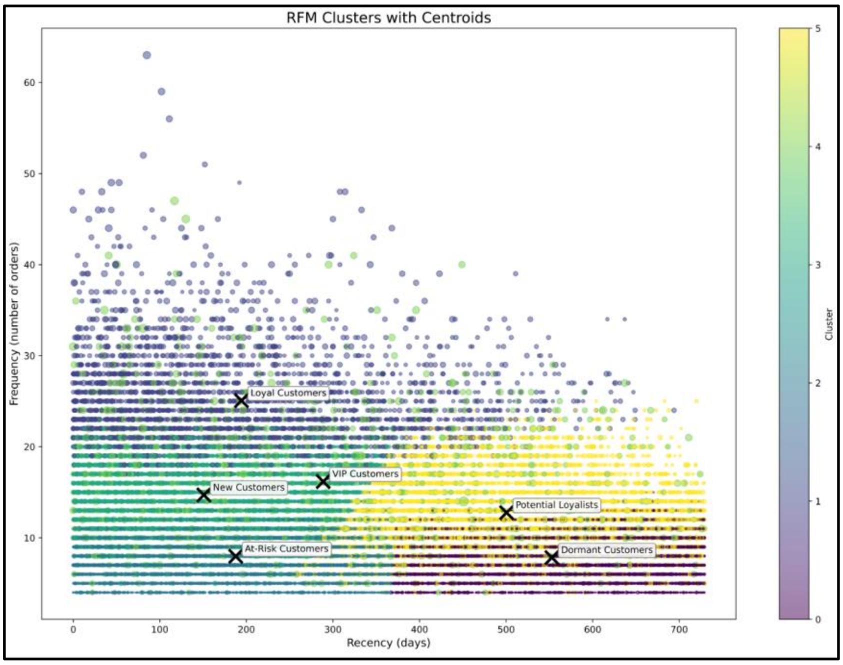

Figure 8. Visualization of Cluster Member Grouping Results

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.14000 p-ISSN : 2541-044X

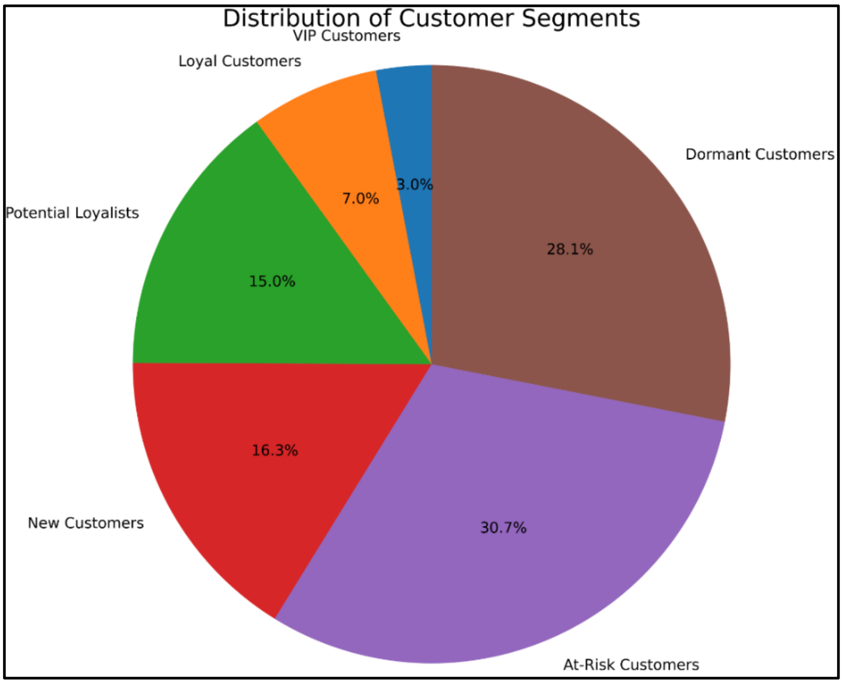

Figure 9. Customer Segmentation Results

The results of customer segmentation displayed in the pie chart provide a visual overview of the distribution of customers into six segments based on RFM (Recency, Frequency, Monetary) analysis. VIP Customers account for 3.0% of total customers, they are the most active customers and have a high spending value, frequent transactions, and make a significant contribution to the company's revenue. Loyal Customers make up 7.0% of total customers, demonstrating strong loyalty with a high frequency of transactions even though the value of their spending may not be as high as VIP Customers. Potential Loyalists account for 15.0% of total customers, indicating the potential to become loyal customers with a fairly high transaction frequency but not yet achieving a very high level of spending. New Customers make up 16.3% of total customers, they have just started transacting with the company and show a fairly good frequency and value of spending, requiring a focus on retaining and increasing their engagement. At-Risk Customers account for 30.7% of total customers, showing signs of declining transaction activity and requiring special attention to prevent them from becoming Dormant Customers. Dormant Customers comprise 28.1% of total customers, they have not transacted for a long time and have a low frequency and value of spending, requiring a strong reactivation strategy to restore their engagement with the company. Furthermore, an evaluation will be carried out as seen in Figure 10.

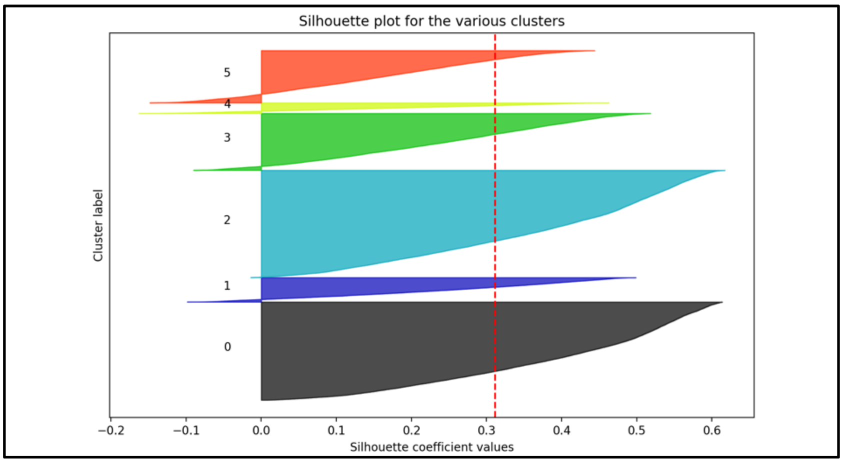

Figure 10. Model Evaluation with the Silhouette Score Method

An average silhouette score of 0.3115 indicates that the clusters formed are quite well defined. This score ranges from -1 to 1, where higher values indicate better clusters. A score of 0.3115 indicates that objects in the cluster tend to be more similar to each other compared to objects in other clusters, although there is still room for improvement.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

This study uses the K-Means clustering method to optimize retail marketing strategies through customer segmentation based on their shopping behavior. RFM (Recency, Frequency, Monetary) analysis results in six customer segments: VIP Customers, Loyal Customers, Potential Loyalists, New Customers, At-Risk Customers, and Dormant Customers. VIP Customers (3.0%) are highly active customers with a high frequency of transactions and significant spending value, so they deserve exclusive offers and premium services. Loyal Customers (7.0%) transact frequently and show strong loyalty, requiring loyalty programs and recurring discounts. Potential Loyalists (15.0%) show higher potential loyalty if given special attention through retention campaigns. New Customers (16.3%) need to focus on brand recognition and product promotion to increase engagement. At-Risk Customers (30.7%) showed a decrease in transaction activity and needed intervention to prevent churn. Dormant Customers (28.1%) have not transacted for a long time and need a strong reactivation strategy. The cluster evaluation showed an average Silhouette score of 03115, which indicates that the clusters formed are quite well defined. This score ranges from -1 to 1, where higher values indicate better clusters. A score of 0.3115 indicates that objects in the cluster tend to be more similar to each other compared to objects in other clusters, although there is still room for improvement. This segmentation allows companies to allocate marketing resources more effectively.

CONCLUSION

This study has shown that customer segmentation using the K-Means clustering method and RFM (Recency, Frequency, Monetary) analysis can provide significant insights to optimize retail marketing strategies. By dividing customers into six segments—VIP Customers, Loyal Customers, Potential Loyalists, New Customers, At-Risk Customers, and Dormant Customers—companies can target each group with a more targeted approach. VIP Customers, who account for 3.0% of total customers, are the most active and valuable, deserving of exclusive offers and premium services. Loyal Customers, at 7.0%, show strong loyalty and frequent transactions, suitable for loyalty programs and special discounts. Potential Loyalists, who comprise 15.0%, have the potential to become more loyal and can increase their engagement through retention campaigns. New Customers, at 16.3%, need a brand recognition and promotion strategy to increase their initial engagement. At-Risk Customers, accounting for 30.7%, showed signs of declining activity and required intervention to prevent churn. Dormant Customers, at 28.1%, have not transacted for a long time and require a strong reactivation strategy. The clustering evaluation showed an average Silhouette score of 0.3115, which indicates that the clusters formed are quite well defined. This value indicates that although clusters tend to be more similar to each other compared to objects in other clusters, there is still room for improvement. This research provides a solid foundation for companies to develop more effective and efficient marketing strategies, as well as increase customer satisfaction and loyalty.

REFERENCES

Abidar, L., Zaidouni, D., Asri, I. E., & Ennouaary, A. (2023). Predicting Customer Segment Changes to Enhance Customer Retention: A Case Study for Online Retail using Machine Learning. *International Journal of Advanced Computer Science and Applications*, *14*(7). https://doi.org/10.14569/IJACSA.2023.0140799

Akhmetbek, S. (2022). FORECASTING CUSTOMER FUTURE BEHAVIOR IN RETAIL BUSINESS USING MACHINE LEARNING MODELS. *Scientific Journal of Astana IT University*, *10*, 45–60. https://doi.org/10.37943/ILMM7870

Asuah, G., & Prikutse, L. F. (2023). Unleashing Customer Insights through K-Means Clustering for Enhanced Retail Decision-Making. *International Journal of Membrane Science and Technology*, *10*(1), 524–531. https://doi.org/10.15379/ijmst.v10i1.2616

Chellaboina, S., Gembali, M., & S, S. Priya. (2022). Product Recommendation based on Customer Segmentation Engine. *2022 2nd International Conference on Intelligent Technologies (CONIT)*, 1–7. https://doi.org/10.1109/CONIT55038.2022.9847990

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), 1251–1257. https://doi.org/10.1016/j.jksuci.2018.09.004

Dewabharata, A. (2022). Customer Segmentation Using the K-Means Clustering as a Strategy to Avoid Overstock in Online Shop Inventory. *Proceedings of the 1st International Conference on Contemporary Risk Studies, ICONIC-RS 2022, 31 March-1 April 2022, South Jakarta, DKI Jakarta, Indonesia*. Proceedings of the 1st International Conference on Contemporary Risk Studies, ICONIC-RS 2022, 31 March-1 April 2022, South Jakarta, DKI Jakarta, Indonesia, DKI Jakarta, Indonesia. https://doi.org/10.4108/eai.31-3-2022.2320688

Gautam, N., & Kumar, N. (2022). Customer segmentation using k-means clustering for developing sustainable marketing strategies. *Business Informatics*, *16*(1), 72–82. https://doi.org/10.17323/2587-814X.2022.1.72.82

Hamdani, N. A., Mutmainah, A., Maulani, G. A. F., Nugraha, S., & Permana, I. (2023). Omnichannel Fashion Retail in Indonesia: How it affects Marketing Performance? *MIX: JURNAL ILMIAH MANAJEMEN*, *13*(1), 49. https://doi.org/10.22441/jurnal_mix.2023.v13i1.004

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.14000
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Husein, A. M., Waruwu, F. K., Batu Bara, Y. M. T., Donpril, M., & Harahap, M. (2021). Clustering Algorithm For Determining Marketing Targets Based Customer Purchase Patterns And Behaviors. *SinkrOn*, *6*(1), 137–143. https://doi.org/10.33395/sinkron.v6i1.11191

John, J. M., Shobayo, O., & Ogunleye, B. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, *2*(4), 809–823. https://doi.org/10.3390/analytics2040042

Joung, J., & Kim, H. (2023). Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews. *International Journal of Information Management*, *70*, 102641. https://doi.org/10.1016/j.ijinfomgt.2023.102641

Koul, S., & Philip, T. M. (2021). Customer Segmentation Techniques on E-Commerce. *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 135–138. https://doi.org/10.1109/ICACITE51222.2021.9404659

Kristanto, F. H., Wimanda Rahma, H., & Nahrowi, M. (2022). Factors Affecting E-Commerce Customer Loyalty In Indonesia. *Jurnal Syntax Transformation*, *3*(09), 1150–1164. https://doi.org/10.46799/jst.v3i09.613

Li, M., Li, M., Cheng, Y., & Ryu, K. H. (2023). A Novel Distributed K-Means Clustering Algorithm for Big Text Data. *Proceedings of the 2023 7th International Conference on High Performance Compilation, Computing and Communications*, 27–33. https://doi.org/10.1145/3606043.3606048

Li, Z., Zhao, N., Zhang, S., Sun, Y., Chen, P., Wen, X., Ma, M., & Pei, D. (2022). *Constructing Large-Scale Real-World Benchmark Datasets for AIOps* (Versi 1). arXiv. https://doi.org/10.48550/ARXIV.2208.03938

Maddumala, V. R., Chaikam, H., Velanati, J. S., Ponnaganti, R., & Enuguri, B. (2022). Customer Segmentation using Machine Learning in Python. *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 1268–1273. https://doi.org/10.1109/ICCES54183.2022.9836018

Mensouri, D., Azmani, A., & Azmani, M. (2022). K-Means Customers Clustering by their RFMT and Score Satisfaction Analysis. *International Journal of Advanced Computer Science and Applications*, *13*(6). https://doi.org/10.14569/IJACSA.2022.0130658

Nie, F., Li, Z., Wang, R., & Li, X. (2023). An Effective and Efficient Algorithm for K-Means Clustering With New Formulation. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 3433–3443. https://doi.org/10.1109/TKDE.2022.3155450

Somasekhar, G., Krishna, K. S., Reddy, A. K., Kumar, T. K., & Somasekhar, G. (2021). Shopper Segmentation Using Multivariate Risk Analysis for Innovative Marketing Strategies: *International Journal of Asian Business and Information Management*, *12*(1), 60–74. https://doi.org/10.4018/IJABIM.20210101.oa4

Tabianan, K., Velu, S., & Ravi, V. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability*, *14*(12), 7243. https://doi.org/10.3390/su14127243

Turkmen, B. (2022). Customer Segmentation With Machine Learning for Online Retail Industry. *The European Journal of Social and Behavioural Sciences*, *31*(2), 111–136. https://doi.org/10.15405/ejsbs.316

Wani, A., Priyanka, M., & Prasath, R. (2023). Unleashing Customer Insights: Segmentation Through Machine Learning. *2023 World Conference on Communication & Computing (WCONF)*, 1–5. https://doi.org/10.1109/WCONF58270.2023.10235136

Xiahou, X., & Harada, Y. (2022). B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(2), 458–475. https://doi.org/10.3390/jtaer17020024

Xin, H., & Zhang, S. (2022). Analysis of e-commerce users' preference behavior based on k-means algorithm. Dalam D. Qiu, X. Ye, & N. Sun (Ed.), *6th International Workshop on Advanced Algorithms and Control Engineering (IWAACE 2022)* (hlm. 106). SPIE. https://doi.org/10.1117/12.2653451

Zaib, R., & Ourabah, O. (2023). Large Scale Data Using K-Means. *Mesopotamian Journal of Big Data*, 36–45. https://doi.org/10.58496/MJBD/2023/006

Zhang, X. (2023). Research on Crocs's Marketing Strategy. *Advances in Economics, Management and Political Sciences*, *4*(1), 278–281. https://doi.org/10.54254/2754-1169/4/20221076

*name of corresponding author