



The Application of Customers Segmentation Using RFM Analysis Method and K-Means Clustering to Improve Marketing Strategy

Salahudin Robo¹, Putri Indah Melani², Patrisia Fernatyanan³, Muh Riandi Widianoro⁴, Sitti Khairul Bariyyah⁵

^{1,2,3,4} Fakultas Ilmu Komputer, Sistem Informasi, Universitas Yapis Papua, Jayapura, Indonesia

⁵ Fakultas Perikanan dan Ilmu Kelautan, Budidaya Perairan, Universitas Yapis Papua, Jayapura, Indonesia

Email: salahudinrobo759@gmail.com¹, putriindahmey217@gmail.com², patrisia210503@gmail.com³, riandipasdu@gmail.com⁴, rhea.tayang@gmail.com⁵

Abstract

This research aims to overcome problems in improving marketing strategies in the Retail Business industry by using effective customer segmentation. The method used is RFM (Recency, Frequency, Monetary) analysis to measure the time proximity, frequency and monetary value of customer transactions, as well as K-Means Clustering to group customers based on their purchasing behavior. The results showed that the combination of these two methods successfully grouped customers into ten different segments, such as "Champions" and "Hibernating," which provided deep insight into customer needs and behavior. The application of this segmentation provides practical benefits in increasing the efficiency of marketing strategies, customer retention and resource optimization. Overall, this research proves that applied customer segmentation techniques can significantly increase customer satisfaction and loyalty, making a valuable contribution to the field of retail marketing.

Keywords: RFM, K-Means, Clustering, Segmentize, Customers.

1. Introduction

Retail business is a vital sector in the economy, where products and services are provided directly to end consumers. Retail businesses cover a wide range of industries, from supermarkets and department stores to clothing outlets and restaurants[1]. In the face of increasingly intense competition and dynamic changes in consumer behavior, retail companies are faced with the challenge of maintaining and increasing their market share [2]. Retail businesses that provide consumers with a wide range of products for everyday needs need to improve their marketing strategies to remain competitive and better meet customer needs. The number of customers and their diverse needs make it important to segment in order to increase the effectiveness and efficiency of marketing strategies.

Customer segmentation in a retail context is the process of grouping customers based on certain characteristics or behaviors to better understand their needs and preferences. Segmentation is crucial because it allows companies to design more optimal and efficient marketing strategies, improve customer experience, and optimize the use of resources. With proper segmentation, businesses can increase customer retention, maximize revenue, and increase brand loyalty [3].

Various approaches that can be applied in customer segmentation are RFM (Recency, Frequency, Monetary)[4], K-Means Clustering [5], Agglomerative Hierarchical Clustering[6], Data Envelopment Analysis[7], Decision Trees[8] and K-Medoids[9]. Agglomerative Hierarchical Clustering allows identifying hierarchical structures in data, but has high time complexity and is less efficient for



large datasets. Data Envelopment Analysis is suitable for evaluating relative efficiency, but is less flexible in handling non-linear data. Decision Trees provide easy-to-understand interpretations, but are susceptible to overfitting. K-Medoids have better resistance to outliers when compared to K-Means, but have higher computing times [9].

To overcome the limitations of each method, an improvement solution that can be taken is to apply data pre-processing techniques to handle outliers. In this case, RFM (Recency, Frequency, Monetary) Analysis and K-Means Clustering were chosen in this research for several reasons. RFM is a simple and easy to implement method, focusing on three main dimensions of relevant customer behavior: proximity to the last transaction, frequency of transactions, and monetary value of transactions [4]. This allows the identification of different customer segments based on their purchasing behavior without requiring very detailed or historical data. However, RFM may not adequately capture the complexity of customer behavior patterns in larger datasets. K-Means Clustering, on the other hand, is an algorithm that is computationally efficient for large datasets, has linear time complexity, and produces groups of clusters with their respective centers, thus facilitating interpretation [10]. The disadvantage is that this algorithm is sensitive to outliers and requires determining the exact number of clusters at the beginning. In addition, validating the clustering results using the Silhouette Coefficient evaluation metric can help ensure the quality of the resulting segmentation [11].

Therefore, the aim of this research is to apply the RFM and K-Means Clustering approaches in customer segmentation analysis, with the main aim of increasing the efficiency of marketing strategies and the quality of customers' shopping experiences. It is hoped that this research will be able to provide a more comprehensive understanding of customer needs and behavior, and support companies in designing more effective and targeted marketing strategies.

2. Research Methodology

The method applied in this research is CRISP-DM, also known as The Cross Industry Standard Process for Data Mining [12], which is a method commonly used in data mining and the most popular.

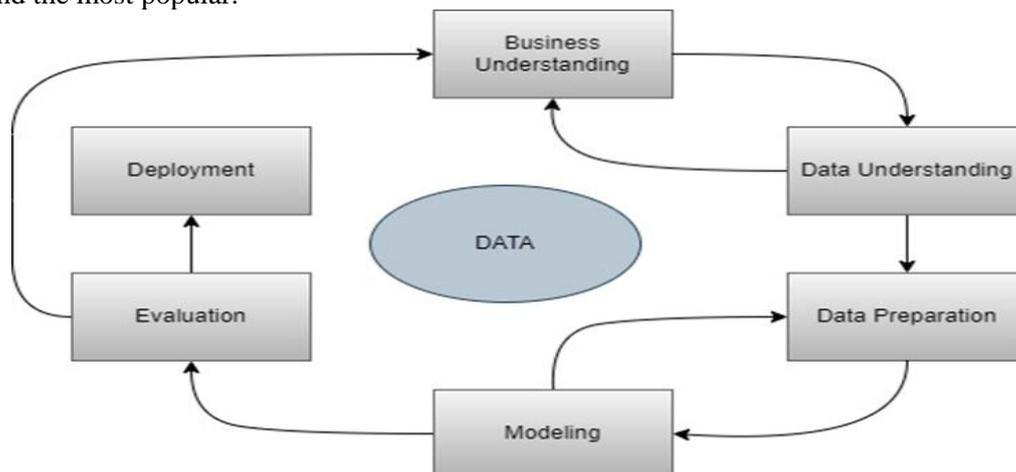
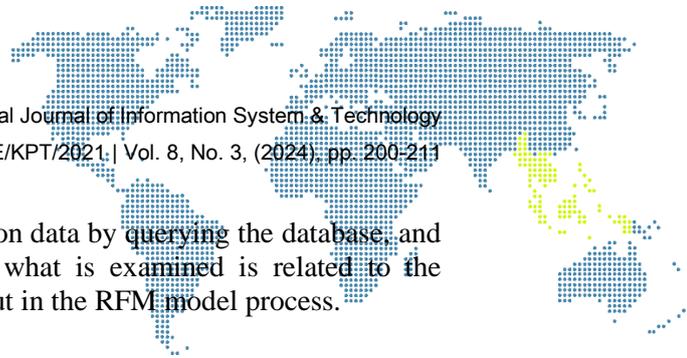


Figure 1. CRISP-DM (The Cross Industry Standard Process for Data Mining)

The following is an explanation of the steps for using the CRISP-DM methodology in this research:

A. Business Understanding

At this stage, data and information are collected and selected from literature studies, reading and studying research related to the research topic, observing research objects,



viewing and understanding outlet shopping transaction data by querying the database, and recording and observing problems systematically. what is examined is related to the research object with the aim of obtaining data as input in the RFM model process.

B. Data Understanding

Establishing the RFM Model uses historical transaction data which functions as a data source for the RFM model. This research uses outlet shopping transaction data for 12 months (January–December) in 2021. The RFM model describes customer consumption behavior based on a database of past transactions in a simplified form into three attributes, namely Recency (R), Frequency (F), and Monetary Value (M)[13]. Recency (R) describes the range of one transaction at a certain time in the past[14]-[15]. The shorter the interval, the greater the R value. Frequency (F) describes the frequency, namely the number of transactions in a certain period. The higher the frequency, the greater the F value[16]. Monetary Value (M) describes monetary value, namely the value of the product in the form of money in a certain period. The greater the amount of money in the period, the higher the value of M[4]. Figure 2 shows the RFM actual value model diagram:

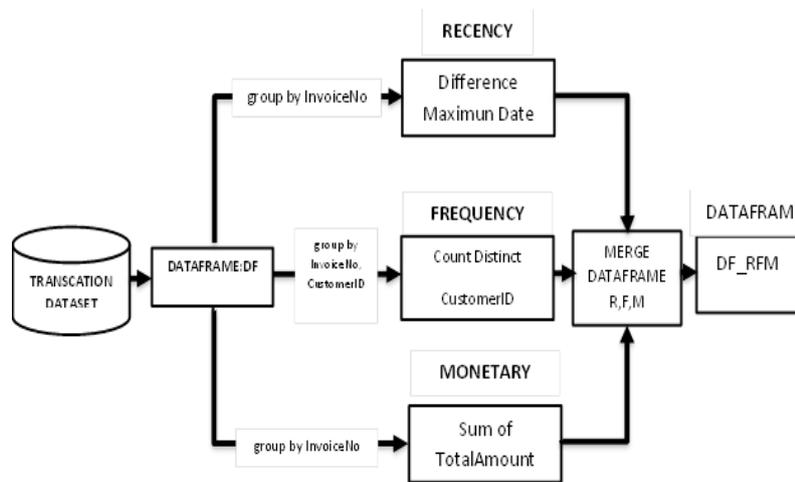


Figure 2. RFM actual value model diagram

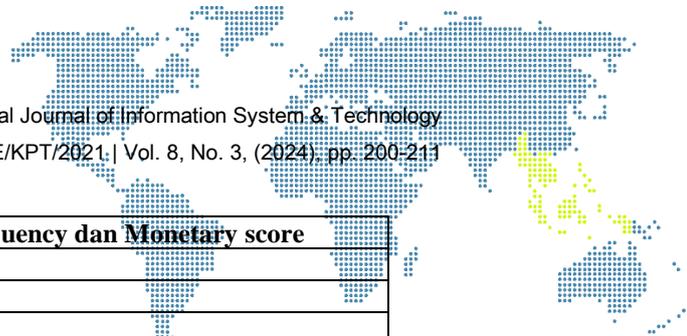
The results of the process of forming a dataset into an RFM model are stored in a data frame with the name `df_rfm`.

C. Data Preparation

Data processing in the RFM model converts RFM values into quantitative scores. This process involves sorting the dataset in descending order based on the R attribute from the newest date to the oldest date, then dividing the dataset into 5 quartiles and assigning a value of 5 to the first 20%, a value of 4 to the second 20%, and so on up to a value of 1[4]. The same process is applied to attributes F and M by sorting F and M in descending order and assigning values. After that, F is sorted in each R category and M is sorted in every combination of R and F categories. This model produces RFM segmentation with the criteria and scores used in RFM analysis as shown in Table 1. In addition, normalization is carried out so that the recency value range , frequency, and monetary are not too different [14].

Table I. RFM Assessment

Criteria	Recency score	Frequency dan Monetary score
Champions	5	4 - 5
Loyal Customers	3-4	4 - 5
Potential Loyalist	4 - 5	2 - 3



Criteria	Recency score	Frequency dan Monetary score
Promosing	4	1
Needs Attention	3	3
About To Sleep	3	1-2
At Risk	1-2	3-4
Can't Lose Them	1 - 2	5
Hibernating	1- 2	1-2
New customers	5	1

Each attribute R, F, and M with actual values will be normalized using standard scaler techniques; the mean becomes the 0 point, and the maximum value is the standard deviation value.

D. Modeling

K-Means is a clustering method that is commonly used in data grouping [17]-[18], including in outlet segmentation analysis based on the RFM model. The basic principle of the K-Means algorithm is very simple. Given a data set X_i ($i = 1, 2, 3, \dots, n$), with predetermined K values and cluster centers [19]. The objective function of the K-Means algorithm is the sum of squared errors (Sum of Squared Errors/SSE) [20].

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} |x - \bar{x}_i|^2 \quad (1)$$

In this research, K-Means is applied to group outlet shopping transaction data into several clusters, with the number of clusters varying from 2 to 8. The process begins by determining the number of clusters which will help identify shopping transaction patterns in outlet segmentation, then determining the initial value of the centroid using random data [21]. Next, the distance between each data point and the centroid is calculated to determine which cluster is closest to each data point. This process is repeated until no objects move clusters, indicating the completion of the clustering process [22].

E. Evaluation

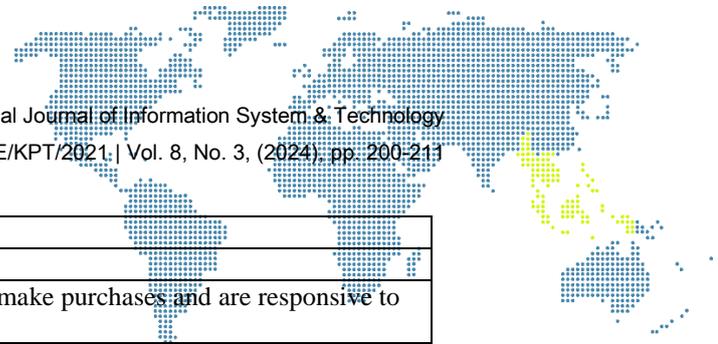
The evaluation stage is carried out using the silhouette index (SI) to measure the quality of the cluster results produced by the K-Means algorithm [17]. The silhouette index provides an overview of how well the data in the cluster is close to each other and how far the cluster is separated from other clusters [10]. It is important to evaluate whether the resulting clusters are sufficiently compact and well separated. The higher the silhouette value, the better the quality of the resulting cluster. Thus, this evaluation helps ensure that outlet segmentation based on the RFM model produces results that are meaningful and can be interpreted well.

F. Deployment

For the deployment stage, analysis is carried out based on the scores given to each retail_id for the recency, frequency and monetary attributes. The score is given on a scale between 5 and 1, where 5 is the highest score, and 1 is the lowest score. In Table 2, the RFM analysis segmentation is displayed with a description for each segmentation criterion. This stage is important for understanding and applying the results of RFM segmentation analysis in the context of business decisions, such as more appropriate marketing strategies and more effective customer relationship management [14].

Table 2. Segment RFM Analysis

Criteria	Description
Champions	Active customers recently did make transactions, buy frequently, and spend



Criteria	Description
	most widely.
Loyal Customers	Customers who consistently make purchases and are responsive to promotions
Potential Loyalist	New customers with average frequency
Promosing	Customers who recently made a purchase but didn't spend a lot of money
Needs Attention	Customers with above average scores for recency, frequency, and monetary
About To Sleep	Customers with below-average frequency and frequency may be hibernating.
At Risk	Customers who shopped some time ago and need to be reactivated
Can't Lose Them	Customers with characteristics in the past often made transactions but currently have not made transactions for a long time
Hibernating	High-frequency, low-value customers tend to become lost customers (inactive customers).

The structured application of CRISP-DM methodology in this research will not only help in solving the problem effectively, but will also provide valuable insights for decision making and continuous improvement.

3. Results And Discussion

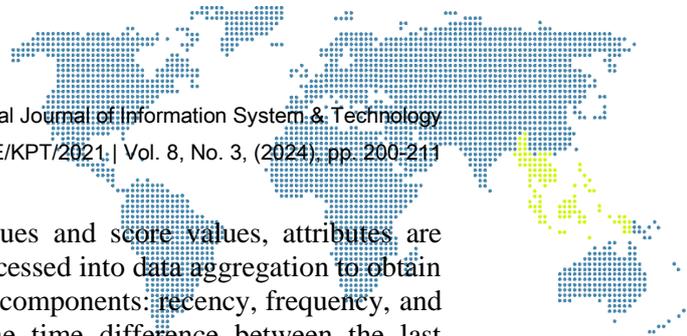
Before data processing is carried out, it is necessary to understand the data to understand the current business. In Figure III, there is a form of shopping transaction dataset, and in Table 3, there is an explanation.

Table 3. Shopping Transaction Data Structure

No	Field Name	Description
1	InvoiceNo	The invoice number is a 6-digit integral number that is unique for each transaction.
2	StockCode	The product code is a 5-digit integral number that is unique for each different item.
3	Description	Product name
4	Quantity	Number of items of each product in one transaction.
5	InvoiceDate	Shows the day and time when each transaction was made
6	UnitPrice	Price per unit of product
7	CustomerID	Customer identification number, is a 5-digit integral number that is unique for each customer.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6.0	12/1/2020 8:26	2.55	17850.0
1	536365	71053	WHITE METAL LANTERN	6.0	12/1/2020 8:26	3.39	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8.0	12/1/2020 8:26	2.75	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6.0	12/1/2020 8:26	3.39	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6.0	12/1/2020 8:26	3.39	17850.0

Figure 3. Shopping Transaction Data Sample



In forming an RFM model based on actual values and score values, attributes are selected for the transaction dataset which is then processed into data aggregation to obtain RFM values. The RFM value consists of three main components: recency, frequency, and monetary. Recency is calculated by measuring the time difference between the last transaction carried out by the outlet and the specified date. Frequency is determined based on the total number of transactions recorded from the outlet within a certain time period. Monetary is calculated based on the total nominal money spent by the outlet to purchase products. Table 4 provides a detailed explanation of the attributes selected for the transaction dataset.

Table 4. Trasaction Data Structure after Attribute Selsection

No	Field Name	Description
1	InvoiceNo	The invoice number is a 6-digit integral number that is unique for each transaction.
2	Quantity	Number of items of each product in one transaction.
3	InvoiceDate	Shows the day and time when each transaction was made
4	UnitPrice	Price per unit of product.

Figure 4 shows example data that includes actual RFM values and RFM scores. The RFM frame data in Figure IV has been normalized using a standard scaler or z-score transformation for each outlet. After the normalization process, the transformed data frame is presented in Figure 5.

	CustomerID	Recency	Frequency	MonetaryValue	Recency_score	Frequency_score	Monetary_score	RFM_Score	rfm_segment
0	12346.0	326	1	77183.600	1	1	5	7	11
1	12747.0	2	88	3489.740	5	4	5	14	54
2	12748.0	1	3724	28515.560	5	5	5	15	55
3	12749.0	4	197	4045.580	5	5	5	15	55
4	12820.0	3	59	942.340	5	4	4	13	54
5	12821.0	214	6	92.720	1	1	1	3	11
6	12822.0	71	46	948.880	2	3	4	9	23
7	12823.0	75	5	1759.500	2	1	4	7	21
8	12824.0	60	24	392.080	3	2	2	7	32
9	12826.0	3	82	1319.720	5	4	4	13	54
10	12827.0	6	25	430.150	5	2	2	9	52

Figure 4. RFM Data Framework Model: Actual Values and Scores

	Recency-saled	Frequency-scaled	MonetaryValue-scaled
0	2.693453	-0.413147	11.472678
1	-0.922272	0.016509	0.268150
2	-0.933432	17.973161	4.073114
3	-0.899953	0.554813	0.352661
4	-0.911112	-0.126710	-0.119161
5	1.443573	-0.388454	-0.248338
6	-0.152257	-0.190911	-0.118166
7	-0.107618	-0.393393	0.005082
8	-0.275013	-0.299560	-0.202823
9	-0.911112	-0.013122	-0.061783
10	-0.877633	-0.294621	-0.197035

Figure 5. RFM Dataframe Normalization

In Figure 5, we can see that the RFM data initially has actual values, but is then normalized using the standard scaler method. Modeling in this research uses the K-Means clustering algorithm and is carried out using the Goggle Colab tool, with the parameters and commands shown in Figure 6.



```
[ ] # Modeling
possible_k = [2, 3, 4, 5, 6, 7, 8, 9, 10]
silhouette = []

for k in possible_k:
    kmeans = KMeans(n_clusters=k, random_state=42).fit(rfm_scaled_df)
    silhouette.append(silhouette_score(X=rfm_scaled_df, labels=kmeans.labels_))
```

Figure 6. K-Means Modeling (Silhouette Coefficient)

The command above is an example of Python code used to perform modeling with the K-Means algorithm and calculate the Silhouette score to find the optimal number of clusters in RFM analysis. The Silhouette Score provides information about how well each sample is clustered, a higher value indicates better cluster quality. In the picture, it can be seen that the number of best clusters is shown in the graph which has the highest percentage in cluster 5 with a value of 0.618023526699823 and the lowest percentage is in cluster 9 with a value of 0.4126166422852428. Based on the Silhouette Coefficient value, the best number of clusters is.

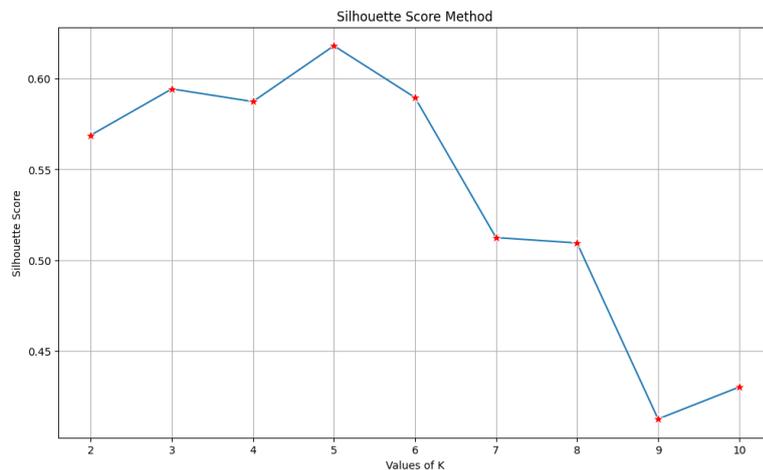


Figure 7. Silhouette Score Results Graph

The K-Means model is run with n_cluster = 5, producing centroid values in the last iteration as shown in figure 8.

Cluster	Cluster Members	Centroid R	Centroid F	Centroid M	
0	0	2749	-0.491698	-0.035830	-0.062395
1	1	3	-0.918552	26.004927	6.353931
2	2	929	1.568887	-0.290528	-0.196883
3	3	2	-0.894373	1.424002	31.984406
4	4	129	-0.785069	2.228951	2.103857

Figure 8. Final Centroid Result

The scatter plot graph in Figure 9 shows how the data is spread out for the monetary recency attribute.

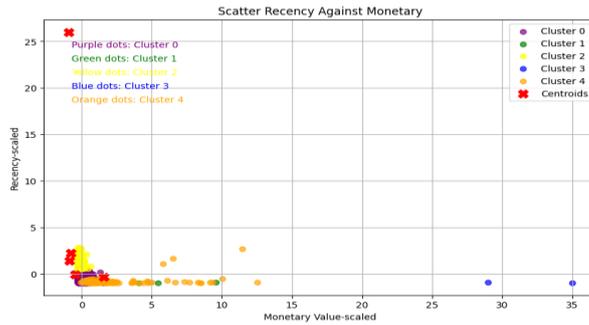


Figure 9. Recency to Monetary Spread

In the picture you can see purple dots, which are data groups with cluster 0, green dots are cluster 1 data points, yellow dots are cluster 2 data points, blue dots are cluster 3 data points and orange dots are cluster 4 data points. Data is distributed by grouping similarities in normalized R and M values. In addition, in Figure 8, it can be seen that each cluster group has a centroid point, which is marked with a red X for each cluster.

The data frame in Figure 10 shows information about outlet grouping based on cluster as well as other additional information, where outlets are also categorized based on segment and score criteria. For example, CustomerID 12346.0 is included in cluster 4 with Hibernating criteria, while 12747.0 is included in cluster 0 with Champions criteria. These results provide additional value to cluster interpretation analysis, which is very useful in a business context.

	CustomerID	Recency	Frequency	MonetaryValue	Recency_score	Frequency_score	Monetary_score	RFM_Score	rfm_segment	Cluster
0	12346.0	326	1	77183.600	1	1	5	7	Hibernating	4
1	12747.0	2	88	3489.740	5	4	5	14	Champions	0
2	12748.0	1	3724	28515.560	5	5	5	15	Champions	1
3	12749.0	4	197	4045.580	5	5	5	15	Champions	0
4	12820.0	3	59	942.340	5	4	4	13	Champions	0
5	12821.0	214	6	92.720	1	1	1	3	Hibernating	2
6	12822.0	71	46	948.880	2	3	4	9	At-Risk	0
7	12823.0	75	5	1759.500	2	1	4	7	Hibernating	0
8	12824.0	60	24	392.080	3	2	2	7	About to Slip	0
9	12826.0	3	82	1319.720	5	4	4	13	Champions	0
10	12827.0	6	25	430.150	5	2	2	9	Potential Loyalists	0

Figure 10. Cluster Results Data with Segments and Scores

Apart from that, this research also provides information about outlet bar graphs based on existing clusters and score criteria, as seen in the graph in Figure 11.

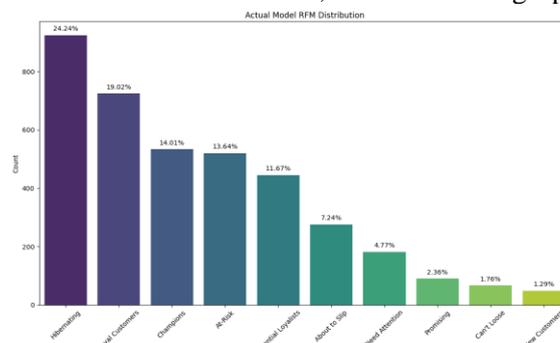
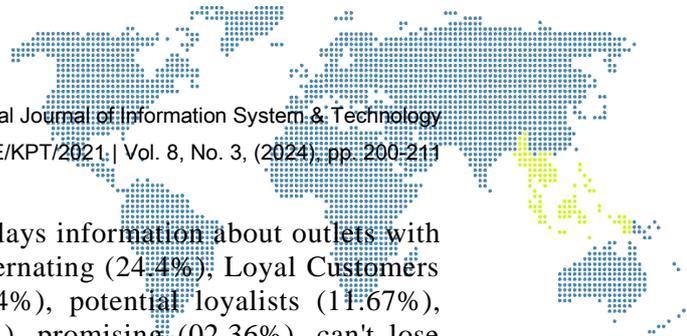


Figure 11. Segment RFM Distribution graph



The outlet segment bar graph in Figure X displays information about outlets with the composition of segment criteria, namely, hibernating (24.4%), Loyal Customers (19.02%), Champions (14.01%), At-Risk (13.64%), potential loyalists (11.67%), About to Sleep (7.24%), Need attention (4.77%), promising (02.36%), can't lose them (1.76%) and New Customers (1.29%).

No	Segment	Cluster	Member	
0	1	Hibernating	4	924
1	2	Loyal Customers	0	725
2	3	Champions	0	534
3	4	At-Risk	0	520
4	5	Potential Loyalists	0	445
5	6	About to Slip	0	276
6	7	Need Attention	0	182
7	8	Promising	0	90
8	9	Can't Loose	0	67
9	10	New Customers	0	49

Figure 12. number of members for each segment

Figure 12 explains that there are ten customer segments with different characteristics. Among them are "Hibernating" with 924 inactive members, "Loyal Customers" with 725 loyal members, and "Champions" with 534 members who are the best customers. There is also "At-Risk" with 520 members who may stop becoming customers, as well as "Potential Loyalists" with 445 members who have the potential to become loyal. Other segments include "About to Slip" with 276 members who need more attention, and "Need Attention" with 182 members who need attention to stay active. Companies can use this information to design more effective marketing and customer retention strategies.

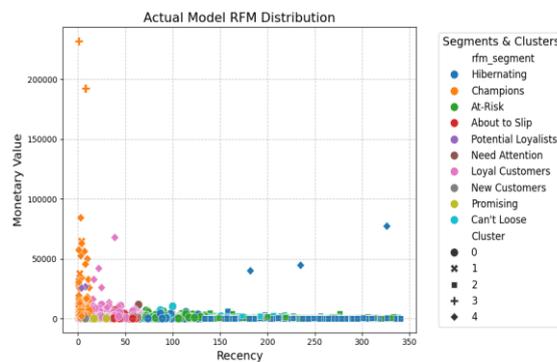
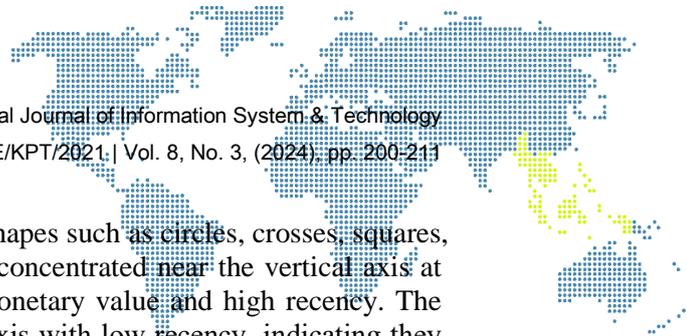


Figure 13. Actual Model RFM Distribution

Figure 13 displays the actual RFM (Recency, Frequency, Monetary) model distribution of various customer segments and clusters. The vertical axis (Monetary Value) shows the amount of money spent by customers, where the higher the position of the point on this axis, the greater the monetary value spent. The horizontal axis (Recency) shows how long it has been since the customer last made a transaction, with points further to the right indicating more time since the last transaction. Each dot on the graph represents a customer, with colors and shapes indicating that RFM segment and customer cluster.

To the right of the graph explains the color codes and shapes for various segments and clusters, such as "Hibernating" indicating inactive customers with low monetary value, and "Champions" indicating frequent transacting customers with high monetary value in



relatively recent times. Clusters are represented by shapes such as circles, crosses, squares, and plus signs. The data distribution appears to be concentrated near the vertical axis at the bottom, indicating many customers with low monetary value and high recency. The “Champions” segment is at the top of the vertical axis with low recency, indicating they have recently made large transactions, while the “Hibernating” segment is spread across the bottom right of the graph, indicating they have not transacted for a long time and have low monetary value.

Using these charts, businesses can identify customer groups that require special attention, formulate more efficient marketing strategies and increase customer loyalty by understanding their behavior based on the RFM model.

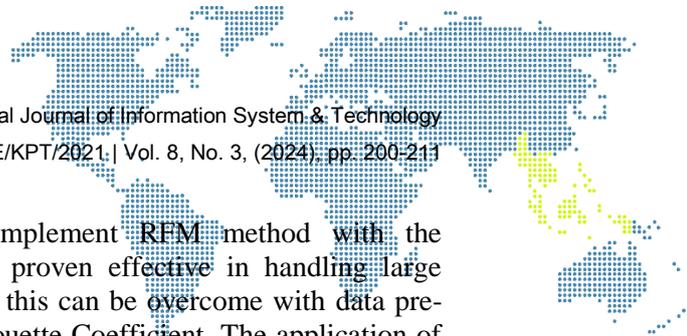
Table 5. Recommendations for Improving Marketing Strategy

Criteria	Recommendations for Improving Marketing Strategy
Champions	Give special rewards and priority access to new products. By providing offers such as exclusive loyalty programs and special discounts.
Loyal Customer	Increase engagement with exclusive offers. By conducting personal email campaigns, product bundling promos, and referral programs.
Potential Loyalist	Convert them into loyal customers. By providing discounts on second purchases, educational content about products, and proactive customer service.
Promising	Encourage them to increase purchases. Through special promos for second purchases, upselling and cross-selling.
Needs Attention	Kembalikan minat mereka sebelum mereka hilang. Dengan penawaran terbatas waktu, pengingat produk yang sering dibeli, dan survei kepuasan.
About To Sleep	Reactivate before they actually stop shopping. By conducting reactivation campaigns with big discounts, personal phone calls, and reminder emails.
At Risk	Identify the reasons they stop shopping and address those issues with satisfaction surveys, return special offers, and dedicated customer service.
Can't Lose Them	Increase their engagement again. With exclusive offers for repurchases, email reminders of favorite products, and VIP service.
Hibernating	Reactivate with attractive offers. By carrying out big discounts, bundling offers, and reminder campaigns via email.

The table above presents recommendations for improving marketing strategies based on RFM segment analysis. Each customer segment has different characteristics, and to optimize customer retention and improve business performance, an approach is needed that is appropriate for each segment. For example, for the “Champions” segment which is active customers with high frequency and high spending value, recommended marketing strategies include providing special rewards and priority access to new products. Meanwhile, for the "At Risk" segment which is customers who have stopped shopping, it is recommended to identify the reasons they stopped shopping and offer special offers to return, as well as provide special customer service. Thus, these recommendations can support directing marketing efforts to be more effective and efficient according to the needs and behavior of each customer segment.

4. Conclusion

The research results show that this method successfully groups customers into ten segments with different characteristics, such as "Champions," "Loyal Customers," and "Hibernating." This segmentation helps in digging a deeper understanding of customer needs and behavior. Based on the segmentation results, this research provides specific marketing strategy recommendations for each segment, for example, providing special rewards and priority access to new products for "Champions" customers, as well as offering attractive offers to encourage reactivation of "Hibernating" customers.



The combination of the simple and easy to implement RFM method with the computationally efficient K-Means Clustering has proven effective in handling large datasets. Although K-Means is sensitive to outliers, this can be overcome with data pre-processing and validation techniques using the Silhouette Coefficient. The application of this segmentation provides practical benefits for Retail Business in increasing the efficiency of marketing strategies, customer retention and resource optimization. Overall, this research makes a significant contribution to the field of retail marketing through the application of appropriate customer segmentation techniques, which can ultimately increase customer satisfaction and loyalty.

References

- [1] S. F. Salem And A. B. Alanadoly, “Driving Customer Engagement And Citizenship Behaviour In Omnichannel Retailing: Evidence From The Fashion Sector,” *Spanish J. Mark. - Esic*, Vol. 28, No. 1, Pp. 98–122, 2024, Doi: 10.1108/Sjme-10-2022-0220.
- [2] K. Hameli, “A Literature Review Of Retailing Sector And Business Retailing Types,” *Ilir. Int. Rev.*, Vol. 8, Jul. 2018, Doi: 10.21113/Iir.V8i1.386.
- [3] J. M. John, O. Shobayo, And B. Ogunleye, “An Exploration Of Clustering Algorithms For Customer Segmentation In The Uk Retail Market,” *Analytics*, Vol. 2, No. 4, Pp. 809–823, 2023, Doi: 10.3390/Analytics2040042.
- [4] Samidi, R. Y. Suladi, And D. Kusumaningsih, “Comparison Of The Rfm Model’s Actual Value And Score Value For Clustering,” *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 7, No. 6, Pp. 1430–1438, 2023, Doi: 10.29207/Resti.V7i6.5416.
- [5] Emigawaty, K. Adi, A. F. Rochim, B. Warsito, And A. Wibowo, “K-Means Clustering Algorithm For Partitioning The Openness Levels Of Open Government Data Portals,” *Int. J. Informatics Vis.*, Vol. 7, No. 3, Pp. 977–983, 2023, Doi: 10.30630/Joiv.7.3.1761.
- [6] R. Ananda And A. Prasetiadi, “Hierarchical And K-Means Clustering In The Line Drawing Data Shape Using Procrustes Analysis,” *Int. J. Informatics Vis.*, Vol. 5, No. 3, Pp. 306–312, 2021, Doi: 10.30630/Joiv.5.3.532.
- [7] M. Izadikhah, R. Farzipoor Saen, K. Ahmadi, And M. Shamsi, “How To Use Fuzzy Screening System And Data Envelopment Analysis For Clustering Sustainable Suppliers? A Case Study In Iran,” *J. Enterp. Inf. Manag.*, Vol. 34, No. 1, Pp. 199–229, 2020, Doi: 10.1108/Jeim-09-2019-0262.
- [8] B. Charbuty And A. Abdulazeez, “Classification Based On Decision Tree Algorithm For Machine Learning,” *J. Appl. Sci. Technol. Trends*, Vol. 2, No. 01, Pp. 20–28, 2021, Doi: 10.38094/Jastt20165.
- [9] E. Wang Yan, A. Jian, L. Yan, And W. Honggang, “Optimization Of K-Medoids Algorithm For Initial Clustering Center,” *J. Phys. Conf. Ser.*, Vol. 1487, No. 1, 2020, Doi: 10.1088/1742-6596/1487/1/012011.
- [10] S. Liu, “An Improved K-Means Clustering Based On Differential Evolution,” *J. Phys. Conf. Ser.*, Vol. 2595, No. 1, Pp. 1–7, 2023, Doi: 10.1088/1742-6596/2595/1/012010.
- [11] M. Shutaywi And N. N. Kachouie, “Silhouette Analysis For Performance Evaluation In Machine Learning With Applications To Clustering,” *Entropy*, Vol. 23, No. 6, Pp. 1–17, 2021, Doi: 10.3390/E23060759.
- [12] C. Schröer, F. Kruse, And J. M. Gómez, “A Systematic Literature Review On Applying Crisp-Dm Process Model,” *Procedia Comput. Sci.*, Vol. 181, No. 2019, Pp. 526–534, 2021, Doi: 10.1016/J.Procs.2021.01.199.
- [13] S. Chavhan, “R F M A N A Ly S I S F O R C U S T O M E R Segme N Tatio N Usi N G Machi N E Lear N I N G : A Survey Of A Decade Of Research,” Vol. 11, No. December, Pp. 166–173, 2022.

- 
- [14] I. Karacan, I. Erdogan, And U. Cebeci, “A Comprehensive Integration Of Rfm Analysis, Cluster Analysis, And Classification For B2b Customer Relationship Management,” *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, Pp. 497–508, 2021.
- [15] A. Gülcü And S. Çalışkan, “Clustering Electricity Market Participants Via Fm Models,” *Intell. Decis. Technol.*, Vol. 14, No. 4, Pp. 481–492, 2020, Doi: 10.3233/Idt-200092.
- [16] I. Huda, A. A. Suhendra, And M. A. Bijaksana, “Design Of Prediction Model Using Data Mining For Segmentation And Classification Customer Churn In E-Commerce Mall In Mall,” *Int. J. Informatics Vis.*, Vol. 7, No. 4, Pp. 2280–2289, 2023, Doi: 10.30630/Joiv.7.4.2414.
- [17] C. Yuan And H. Yang, “Research On K-Value Selection Method Of K-Means Clustering Algorithm,” *J.*, Vol. 2, No. 2, Pp. 226–235, 2019, Doi: 10.3390/J2020016.
- [18] M. Harahap, A. W. D. R. Zamili, M. A. Arvansyah, E. F. Saragih, S. Rajen, And A. M. Husein, “K-Means Clustering Algorithm Approach In Clustering Data On Cocoa Production Results In The Sumatra Region,” *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 6, No. 6, Pp. 905–910, 2022, Doi: 10.29207/Resti.V6i6.4199.
- [19] V. H. Antonius And D. Fitrihanah, “Enhancing Customer Segmentation Insights By Using Rfm + Discount Proportion Model With Clustering Algorithms,” *Int. J. Adv. Comput. Sci. Appl.*, Vol. 15, No. 3, Pp. 902–911, 2024, Doi: 10.14569/Ijacs.2024.0150390.
- [20] W. Chang *Et Al.*, “Analysis Of University Students’ Behavior Based On A Fusion K-Means Clustering Algorithm,” *Appl. Sci.*, Vol. 10, No. 18, 2020, Doi: 10.3390/App10186566.
- [21] A. Sharma, S. Vishwakarma, And A. K. Dubey, “Enhancing Data Analysis Through K-Means With Foggy Centroid Selection,” *Int. J. Adv. Comput. Res.*, Vol. 13, No. 64, Pp. 55–62, 2023, Doi: 10.19101/Ijacr.2023.1362018.
- [22] Z. Rahman, M. S. Hossain, M. Hasan, And A. Imteaj, “An Enhanced Method Of Initial Cluster Center Selection For K-Means Algorithm,” *Proc. - 2021 Innov. Intell. Syst. Appl. Conf. Asyu 2021*, 2021, Doi: 10.1109/Asyu52992.2021.9599017.