Elinvo (Electronics, Informatics, and Vocational Education)
Vol 9, No. 1, 2024, pp. 24-37

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

A Machine Learning Model for Local Market Prediction Using RFM Model

Muhammad Yahya[1], Jumadi Mabe Parenreng[1*], Fathahillah[1], Abdul Wahid[1], M. Syahid Nur Wahid[1], Muhammad Fajar B[1]

[1]Universitas Negeri Makassar, Makassar, Indonesia

Article Info

Article history:
Received Dec 29, 2023
Revised May 7, 2024
Accepted May 8, 2024

Keywords:
RFM; K-Means; data mining; e-commerce

Abstract

This study explores the application of machine learning for local market prediction in e-commerce. By leveraging the RFM segmentation method, the model predicts product sales based on user shopping patterns. The RFM score, calculated using recency, frequency, and monetary values of customer purchases, segments customers into distinct categories. The research utilizes a dataset obtained through seven parameters and performs data preprocessing. K-Means clustering then classifies customers into Low, Medium, and High levels based on their RFM scores. Customers in the Low category exhibit low purchase activity but high product browsing. The Medium segment displays consistent purchases of a limited product range. High-level customers demonstrate frequent purchases with significant spending. The identified customer segments enable targeted marketing strategies. For Low-level customers, discounts or product feature promotions can incentivize purchases. Combining product offerings can entice Medium-level customers to explore new products. Finally, High-level customers can be engaged through loyalty programs offering rewards. This approach empowers e-commerce sellers to tailor marketing strategies for each customer segment, enhancing market dominance.

Corresponding Author:
Email: jparenreng@unm.ac.id

INTRODUCTION

The current trend of the market space is one of the important types of information for sellers to increase profits in selling a product. In this era, the consumer or buyer every day is under economic, social, political, cultural and technological influence. Today, the internet is used as a means to find information about products and make transactions. Therefore, it has become an integral part of daily life for millions of users and consumers around the world [1]. With the advent of the internet and the development of digital services, the online shopping environment is also evolving, sellers can offer consumers more product choices in the buying process, providing them with better services and products [2]. The development of digital services has also made e-commerce grow rapidly offering many opportunities for businesses to develop and be more profitable for sellers and make it easier for consumers in the process of purchasing a product [3]. The rapid development of e-commerce brings problems for companies or sellers who are trying to develop strategies in e-commerce. This is very difficult given the flow of information technology and applications. There seems to be a constant stream of new software. However, companies or sellers who promote their e-commerce strategy to some extent worry that they will lose customers to competitors if they do not have an e-commerce strategy [1], [4].

But nowadays, sellers still find it difficult to find out the trend of market products because the market trends change every day. This is because the purchase of products by each consumer is different according to the trends that occur in the surrounding environment. Consumers are also stimulated through different psychosocial traits such as income, buying motivation, company presentation, company or brand presence in social networks, demographic variables (age, gender, disposable income, etc.), workplace, payment approach, type of store (online or physical), etc. [5], [6], [7], [8]. According to Assurance, A trend will be formed if one of the following criteria is met, namely: 1)Seven data appears repeatedly all above or all below the statistical average. 2)Seven recurring upgrade data. 3)Seven repetitive data were reduced. 4)Ten out of eleven data appear repeatedly, all above or all below the statistical average. 5)Two of the three consecutive points are outside the two standard deviations above or below the statistical average. 6)Four out of five repetitive data points are outside one standard deviation above or below the statistical average. Machine learning aims to create algorithms that can be studied and generate statistical models for data analysis and prediction. Machine learning algorithms can self-study based on existing or provided data and produce accurate predictions without having to be programmed specifically for a specific task. Apart from theoretical developments, recent years have seen rapid developments in the application of machine learning, some of which are the development of AI algorithms and other researchers in various fields who are adopting machine learning for their own purposes [9]. Customer segmentation plays a crucial role in e-commerce by enabling sellers to tailor their marketing strategies to specific customer groups. The RFM (Recency, Frequency, Monetary) model is a well-established segmentation technique that utilizes customer purchase history data to categorize them based on their recent purchase activity, purchase frequency, and monetary value of purchases [10]. Machine learning algorithms, such as K-Means clustering, can automate the segmentation process, leading to more efficient and objective customer grouping [11]. Several studies have explored the effectiveness of targeted marketing strategies for different customer segments in e-commerce [12], [13]. This research builds upon this existing knowledge by leveraging RFM segmentation and K-Means clustering to identify customer segments in a local e-commerce platform specifically designed for villagers. We then explore the potential of this segmentation for developing targeted marketing strategies tailored to each customer group.

Research to determine the results of analysis and prediction of sales or product data has been carried out previously using different methods and parameters such as KNN implementation for prediction of skincare packaging sales; in this study, predictions were carried out on skincare sales that are most in demand. To find out the sales of skincare packaging products that are most in demand, data mining classification techniques are carried out using the K-Nearest Neighbor algorithm. The results of the accuracy calculation test to find out the sales in the next few months obtained the results of the 80% accuracy value [14]. However, the study took one parameter, namely sales results. Our segmentation aligns with previous research by [15], who found that discounts can be effective in incentivizing purchases from low-engagement customers. Similarly, [16] demonstrated the success of product bundling in attracting medium-engagement customers to explore new product categories. Furthermore, our identification of a high-engagement customer segment aligns with studies by [17] who advocate for loyalty programs to retain these valuable customers. This research explores the application of machine learning models in conjunction with RFM (Recency, Frequency, Monetary) segmentation to achieve three key objectives in the local e-commerce market. Firstly, by leveraging RFM analysis, we aim to segment customers into distinct categories based on their purchasing behavior. Secondly, we utilize this segmentation alongside a machine learning model to predict product sales for different customer segments. Finally, by understanding these segments, we seek to develop targeted marketing strategies that can be tailored to each group (Low, Medium, High) based on their unique buying patterns, ultimately enhancing market dominance for e-commerce sellers. Predict market trends using machine learning: a) Market Trends, market trends indicate the direction of movement of the market. There are only two directions where the market can be an uptrend or a downtrend [18], [19] there will also be a

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

phase where there is no trend and the price moves mainly sideways for some time. Such periods without trends are often referred to as consolidations, or trading ranges. b) Market Trend Prediction, social platforms are dynamic and can be considered as a new type of source of information for predicting future trends with the application of data analysis techniques [20] Now we can predict market trends using machine learning. Information scientists have begun to solve the problem of prediction with the advancement of learning techniques. In addition, computer scientists have all begun to use machine learning strategies to improve the performance of prediction models and improve prediction accuracy [19], [21]. c) Target Market, the target market is a set of buyers who share the same needs or characteristics that the company decides to serve. The target market evaluates each segment's interest in the market and selects one or more segments to enter.

There are several types of target market strategies, namely: 1) Undifferentiated marketing (or mass marketing), using this strategy, the company decided to ignore the differences in market segments and fill the entire market with a single offer. That is, the strategy is more focused on the needs of consumers in general than others. 2) Differentiated marketing, this strategy is used by companies to target several market segments and design separate bids to each market segment. By offering a wide variety of products and marketing into segments, the company hopes for higher sales and a stronger position within each market segment. 3) Concentrated (niche) marketing, a strategy that only focuses on marketing its products to one or several groups of buyers, so that product marketing is only aimed at the most potential group of buyers. By focusing on a specific group, companies strive to provide the best products for their target market. In addition, companies are more cost-effective in both production, distribution, and promotion, because everything only focuses on one or two groups. 4) Micromarketing, in this strategy, the company produces products to adjust individual specific (individual marketing) and certain locations (local marketing) [22]: a) Machine Learning, machine learning is a subset of artificial intelligence that uses algorithms and statistical models for computers in performing specific tasks without human interaction [20]. Machine Learning is used to teach machines how to handle extra data efficiently. Sometimes after searching for facts, we cannot interpret statistical extracts from information. If so, we apply machine learning [9], [23]. b) Data Mining, data mining is a process that employs one or more machine learning techniques to analyze and extract knowledge automatically.

Other definitions include induction-based learning, which is the process of forming general concept definitions which is carried out by observing specific examples of the concepts to be studied. Knowledge Discovery in Databases (KDD) is the application of scientific methods to data mining. In this context, data mining is one step of the KDD process [24]. c)RFM Model, the RFM (Recency, Frequency and Monetary) model has been widely applied in several fields, especially in the world of marketing. By adopting the RFM model, a decision maker can effectively identify valuable customers and will be used as the development of an effective marketing strategy. The RFM model is often used for market segmentation. RFM maintains information about the most recent recency, the number of times a customer made a purchase (frequency), and the average money spent (monetary). Customers who have purchased recently, most often, and spent a significant amount of money have the opportunity to react to promotions at the upcoming time [25]. K-Means is one of the most popular clustering methods, due to the simplicity of the algorithm and the speed of selection of the cluster center (centroid). The k-Means method often applies the Euclidean distance formula to iteratively determine the similarity of data in a cluster. The steps of clustering data using the k-Means method can be done by: 1) Determining the number of clusters k. 2) Initialize the value of k as a cluster center (centroid) at random. 3) Group each data into a nearby cluster. The proximity of two data is calculated using the Euclidean distance. 4)Recalculate each centroid by calculating the average of all centroid data with the current cluster member. 5)Regrouping each record (back to step 3) using all new centroids until all centroids are not changed anymore. 6)If the centroid does not change anymore, the clustering process is complete. One of the main problems of the k-Means method is how to determine the optimal number of clusters k [26].

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

METHODS

In conducting this research, there are several methods used to reach the stage that can predict market products that are interested in many consumers, including data acquisition or dataset collection in the form of sales data such as purchase time, quantity of purchase products, to time information, the next method is carried out a segmentation process to determine the score or value of the data obtained, then clustering is carried out to determine the customer level (low, high, medium). The research method is shown in Figure 1.

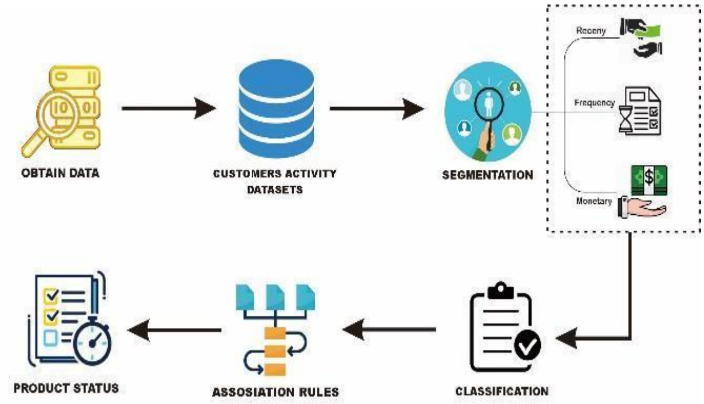

Figure 1. Research Methods

The following are the steps in implementing research methods: a) Data Acquisition, at the data collection stage, the data acquisition method is carried out, the source of data obtained from the E-Commerce System that has been created, the data includes the entire customer activity starting in conducting product searches, product transactions and so on [27]. The data parameters used for later execution are Order_id that store the amount of transaction data, product_id, quantity or number of goods purchased, unit_price or price of per item, Date, and CustomerID. The data source used comes from the results of using the e-commerce web that has been created using the website (https://commerce.laguruda.id/) so that transaction data from buyers to sales data from sellers can be obtained for analysis. The data received is the result of sales or purchases for the last 5 months. b) Segmentation, the segmentation process, the RFM **method, is used** to analyze and group patterns of consumer activities [28]. This method specifically evaluates how long they make a purchase, how often they make a purchase and how much money they spend. The concept of segmentation in RFM analysis will be improved to be more objective and accurate with a clustering approach using the k-Means method, so that the clusters to be formed have optimal data similarity. This can make determining the number of clusters and dataset intervals for each cluster more qualified and precise. Here is the calculation of the RFM score.

$$RFM\ score\ = (Recency\ score\ x\ Recency\ weight) + (Frequency\ score\ x\ Frequency\ weight) + (Monetary\ score\ x\ Monetary\ weight) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

The algorithm implementation is using K Means Clustering with the following steps, K-Means Clustering, the RFM model assigns scores based on recency, frequency, and monetary value of purchases, but manually segmenting customers into distinct groups can be laborious. K-Means clustering addresses this by automatically grouping customers with similar RFM scores into predefined segments (Low, Medium, High in your case). It iteratively calculates distances between customer data points (RFM scores) and cluster centers, assigning them to the closest cluster and recalculating the centers until a stable configuration is achieved. This automation ensures objective and efficient segmentation, especially for large datasets. K-Means algorithm is proposed to determine the cluster of each data that has been segmented. The specified cluster is divided into three, namely *low*, *medium*, and *high.* Customer segmentation will help analyze customer composition accurately and advance the quality of service and marketing.

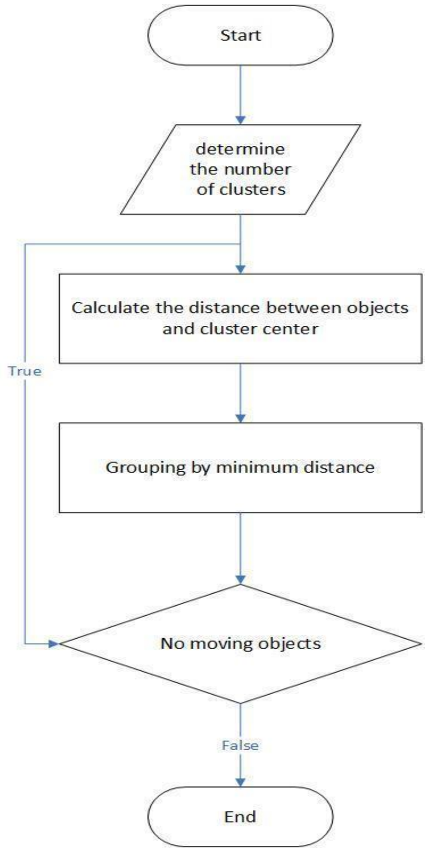

Figure 2. K-Means Algorithm Flowchart

Based on the flowchart in Figure 2, the computational process for K-Means is described as follows: 1) Randomly initialize vector centroid cluster. 2) For every data in the vector, calculate the distance between the vector data with each centroid cluster to determine the minimum data vector within the cluster and the distance calculated using the equation:

$$d(Zp, Mj) = \sqrt{\sum_{k=1}^{d} \quad (Zp, k - Mj, k)} \text{ ....} \quad (2)$$

Where Zp is the *p* point of data, Mj is the *centroid* of the data's *j* cluster. 4) Calculate *centroid* cluster using the equation:

$$M_j = 1/n_j(\sum \quad Z_p) \text{ ....} \quad (3)$$

where n_1 is the amount of data points in cluster j. 5) Repeat step 3 and 4 until it stops meeting the criteria. A satisfactory criterion can be the number of iterations or a change in the position of centroids in sequential iterations.

RESULT AND DISCUSSION

This research focuses on the RFM segmentation work process with the K-Means *Machine Learning* model, which is focused on activities in the E-Commerce System that have previously been created. This E-Commerce system is intended for villagers who have commodities or businesses that they will sell on our platform. From the results of using this system which was carried out for 3 months, we obtained a number of data where the Segmentation process will be carried out on customers. This segmentation process is carried out to find out the behavior of customers in interacting to find the product they want. By using the RFM model we can find out their interest in a product, how often they make purchase transactions and how much money they spend in buying a product. These 3 parameters will later be processed to find out the number of scores owned and then will be clustered using the K-Means algorithm.

In the E-Commerce Application created, there are several categories of product sales, ranging from garden products, seafood, handicrafts, and design results. Each of these categories has different types of products. In this study we took 3 categories to carry out testing, namely Marine Products, Livestock Products, and Handicraft Products. In the initial stage, data normalization is carried out to rid off inappropriate values, then the data is selected based on sales for shrimp products. The data displayed can be seen in the Figure 3.

	order_id	product_id	product	quantity	price	tgl	CustomerID	TotalAmount
0	536365	24	Udang	7	100000	2022-04-03	17850	700000
1	536370	14	Udang	5	10000	2022-03-28	12583	50000
2	536370	3	Udang	2	25000	2022-04-04	12583	50000
3	536370	13	Udang	2	15000	2022-04-15	12583	30000
4	536372	10	Udang	2	60000	2022-04-18	17850	120000

Figure 3. Shrimp Sales Data

Furthermore, calculations are made to find out the Recency value by taking the value from the date / month / year (labeled by tgl) of the transaction, then a calculation is made to find the Frequency by taking the value of the order_id amount, then the last one is done to find the Monetary value by taking the value of the price or price of the product.

$$RFM = X(Y.\lambda(Latest - X_{max}).\lambda Z \quad .... \quad (4)$$

Where:

X = CustomerID
Y = tgl
Latest = Latest date of transaction from tgl
X_{max} = Max value of order
Z = order_id

In equation (4) calculations are carried out to determine the score value of each parameter, where the value (X) is multiplied by (Y) by taking the data of the last transaction date and then subtracting by (X_{max}) then multiplied by the number of orders (Z). The value obtained for RFM can be seen in the Figure 4.

	CustomerID	Recency	Frequency	Monetary
0	12583	198	3	130000
1	17850	183	8	1457000
2	17854	209	1	320000
3	17878	155	1	50000
4	17884	209	1	42000

Figure 4. RFM Value

Next, labeling is carried out for each RFM score as shown in Figure 5, later products that have low or high scores will be grouped in the RFM Group so that later the clustering process can be carried out clearly.

	CustomerID	Recency	Frequency	Monetary	R	F	M	RFMGroup	RFMScore	RFM_Loyalty_Level
0	12583	198	3	31.70	4	1	1	411	6	Platinum
1	17850	183	8	67.71	4	1	1	411	6	Platinum
2	17854	209	1	10.17	4	4	2	442	10	Silver
3	17878	155	1	15.00	3	4	1	341	8	Platinum
4	17884	209	1	14.75	4	4	2	442	10	Silver

Figure 5. Product labeling based on RFM score

Furthermore, a comparison process is carried out for each parameter, the purpose is to find out the relationship between consumer behavior. The comparison determines how much chance the product has to become a popular product among consumers.

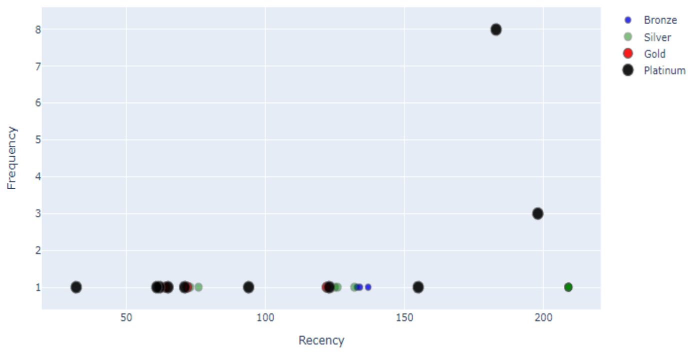

Figure 6. Comparison graph for Frequency and Recency

Figure 6 illustrates the optimal number of clusters determined by the silhouette method, a metric used to measure how similar each point in a cluster is to the points in its own cluster compared to those in other clusters. The graph shows a clear peak at three clusters, indicating that this number provides the best separation among the data points based on their RFM scores. The silhouette score for three clusters is the highest, suggesting that the customers in each cluster are more similar to each other than to those in other clusters. This finding is critical as it validates the optimal cluster number found using the K-means clustering algorithm, ensuring that the chosen clusters are both meaningful and well-separated. It implies that customer segmentation based on three clusters will be the most effective in tailoring marketing strategies to distinct customer groups, leading to more personalized and effective engagement efforts.

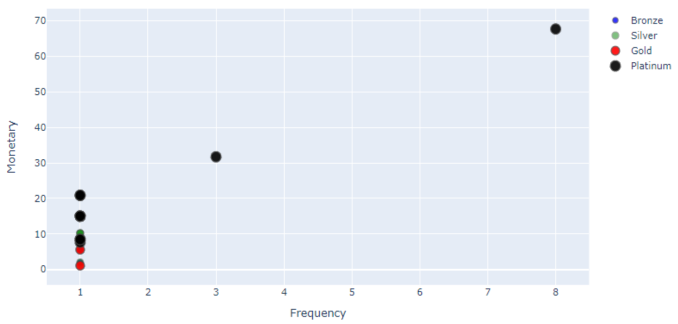

Figure 7. Comparison graph for Monetary and Frequency

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

Figure 7 presents the distribution of customer clusters across different product types. The bar graph categorizes the customers into three clusters and shows the proportion of each cluster within marine products, livestock products, and handicrafts. This distribution highlights significant differences in customer behavior across product categories. For example, the marine products category has a larger proportion of customers in the high-value cluster, indicating a segment of loyal and high-spending customers. In contrast, the handicraft category shows a more even distribution across all three clusters, suggesting a diverse range of customer engagement levels. Understanding these variations helps in designing category-specific marketing strategies. For instance, targeted loyalty programs can be developed for marine products to further engage high-value customers, while more general promotional campaigns may be appropriate for handicrafts to cater to a broader customer base.

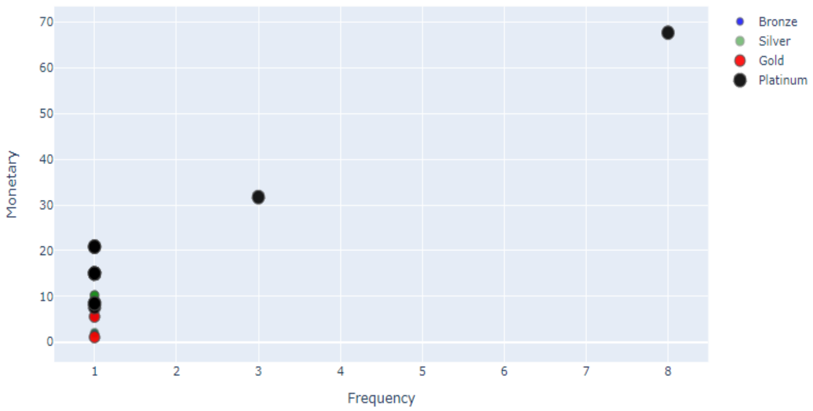

Figure 8. Comparison graph for Monetary and Recency

Figure 8 visualizes the clustering results using a 3D scatter plot where each axis represents one of the RFM dimensions: Recency, Frequency, and Monetary value. The plot clearly shows three distinct clusters, each representing a unique customer segment based on their purchasing behavior. The cluster with low recency and high frequency and monetary value indicates the presence of highly engaged customers who make regular, high-value purchases. Another cluster with high recency and low frequency and monetary value suggests occasional buyers who might need incentives to increase their engagement. The third cluster, positioned between the other two, represents customers with moderate engagement across all dimensions. This visualization is crucial for understanding the dynamics of customer behavior, enabling businesses to identify key segments and develop tailored strategies that cater to the specific needs and preferences of each group. It highlights the effectiveness of the clustering approach in segmenting customers into meaningful groups that can drive targeted marketing efforts and improve customer retention.

Furthermore, the clustering process is carried out using the K-Means algorithm, data that has previously been labeled based on RFM scores. To obtain the mean value, the following equation is used.

$$d(Zp, Mj) = \sqrt{\sum_{k=1}^{d} \quad (Zp, k - Mj, k)}$$

$$M_j = 1/n_j\left(\sum \quad Z_p\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

In equation 5 it can be seen where Zp is the p-th data point, Mj is the *centroid* of the *j*-th data cluster. n_1 is the number of data points in cluster J. To anchor the optimal value, give a distance between 1 and 5 to find out the number of clusters that will later be executed. Consider the chart in Figure 9.

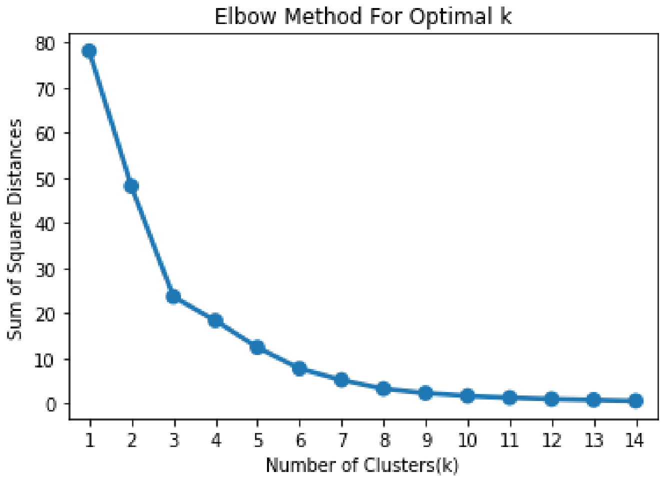

Figure 9. Square Distances Optimal K

Figure 9 displays the optimal value for retrieving the number of clusters, on the graph it decreases squared exactly diagonally by 20, so the cluster taken is 3. In the test results, we took several data samples to test based on CustomerID. To determine the customer level in the clustering of the RFM score results that have been obtained, the level is divided into 3, namely High, Low, and Medium. Customers with a Low level are customers with low purchase order activity, but the activity in tracking products is very high in the long term, then at the Medium level. These customers are the group with the most purchasing activity, but are consistent on only a few products in the same period or period. At the High level, this group of customers is the group that makes the most product purchases with great effort or the costs they spend to buy a product are very much. Before conducting a cluster to determine the customer level, a segmentation process with RFM is carried out as shown in the following table.

Table 1. Marine Products RFM

Customer ID	Recency	Frequency	Monetary
1	173	4	1280000
2	203	1	3000
3	79	5	712000
6	48	3	271000
10	77	5	1005000

Table 1 presents the RFM (Recency, Frequency, Monetary) scores for customers who purchased marine products. Recency indicates the number of days since the last purchase, Frequency represents the total number of transactions, and Monetary measures the total amount spent. The table highlights significant variations among customers, with some showing recent high-frequency purchases while others have infrequent transactions over a long period. This disparity suggests diverse customer behaviors: those with higher Frequency and Monetary values are likely to be loyal, high-value customers who make regular purchases, whereas customers with lower scores might need targeted incentives to increase their purchase frequency and spending. Understanding these patterns allows for the development of tailored marketing strategies to maximize engagement and sales within the marine products segment.

Table 2. Livestock Products RFM

Customer ID	Recency	Frequency	Monetary
1	113	3	525000
2	161	1	35000
3	82	4	4600000
10	112	2	100000
13	57	1	350000

Table 2 details the RFM analysis for livestock products, showcasing how recency, frequency, and monetary value metrics vary across customers. Notably, this table reveals a distinct segmentation where certain customers exhibit very recent and frequent purchases, reflecting a high level of engagement and potential loyalty to livestock products. On the other hand, customers with low frequency and monetary scores, combined with higher recency values, indicate that while they may have recently engaged with the product category, their overall purchasing behavior is limited. This data underscores the potential for targeted promotions and incentives to convert occasional buyers into regular customers, thereby enhancing overall sales within the livestock segment.

Table 3. Handicraft Products RFM

Customer ID	Recency	Frequency	Monetary
1	214	1	80000
2	215	1	35000
10	275	1	250000
18	92	1	2400000
23	276	1	4800000

Table 3 focuses on RFM scores for customers in the handicraft product category. This table reveals diverse customer behaviors, with some customers showing high recency and frequency, indicating they are recent and regular buyers. However, there is a notable segment with low monetary scores despite frequent transactions, suggesting these customers are making smaller purchases. Conversely, high monetary values coupled with low frequency indicate a tendency for bulk purchasing at less frequent intervals. These insights are crucial for devising differentiated marketing strategies, such as offering discounts or bundle deals for high-frequency, low-monetary customers, and personalized loyalty programs for high-value, infrequent buyers, aimed at increasing overall customer lifetime value in the handicraft segment.

The data that has been segmented is then clustered to determine the customer level using the K-Means algorithm. Take a look at the following figure and table.

Marine Products

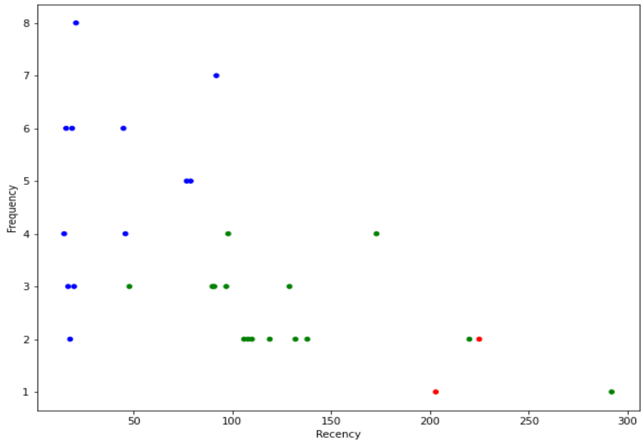

Figure 10. Marine Products Spread Diagram

Figure 10 depicts a scatter plot showing the relationship between recency (days since last purchase) and frequency (number of transactions) specifically for marine products. In this plot, each point represents a customer's purchase history. The general trend observed is that customers with lower recency (recent purchases) exhibit higher frequency of transactions, indicating that they are actively engaged with marine products. Conversely, customers with higher recency values (longer time since last purchase) tend to have lower transaction frequencies, suggesting a decline in engagement. This scatter plot helps identify key customer segments, such as those who are highly engaged and could be targeted for loyalty programs, and those who have not purchased recently and may need re-engagement strategies. Understanding these patterns allows businesses to develop targeted marketing campaigns that aim to retain frequent buyers and reactivate those with declining purchase activity.

Table 4. Customer Level Marine Products

Customer ID	RFM Score	Level	Cluster	Color
1	7	High	1	Green
2	12	Low	0	Red
3	4	Medium	2	Blue
6	8	High	1	Green
10	4	Medium	2	Blue

Table 4 provides a comparative analysis of the frequency and recency of purchases across different product categories. The table illustrates that high-frequency customers tend to have lower recency values, indicating consistent purchasing behavior, whereas those with low frequency generally exhibit higher recency, suggesting a more sporadic buying pattern. This comparison highlights the need for category-specific marketing interventions; for example, frequent but recent buyers may benefit from loyalty rewards to maintain their engagement, while those with higher recency and low frequency might be targeted with promotional campaigns to incentivize more regular purchases.

Livestock Products

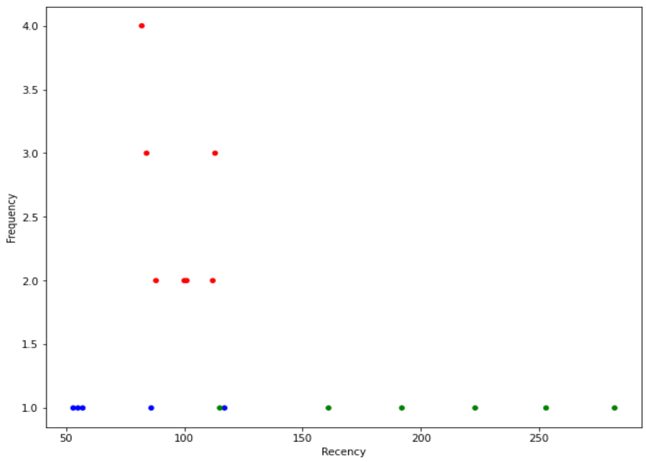

Figure 11. Livestock Spread Diagram

Figure 11 presents a scatter plot illustrating the relationship between recency and frequency for customers who purchase livestock products. The plot shows a similar trend to Figure 10, where customers with low recency values tend to have higher frequencies of purchases, indicating they are actively buying livestock products. Clusters of points can be seen, indicating distinct groups of customers with varying levels of engagement. Customers with both high recency and low frequency are at risk of churn and may benefit from targeted marketing efforts to encourage repeat purchases. This scatter plot is critical for identifying which customers are frequent buyers and which ones are less engaged, providing insights into how to tailor marketing strategies to different customer segments. By focusing on maintaining engagement with frequent buyers and reactivating those with higher recency, businesses can optimize their customer retention strategies and enhance overall sales.

Table 5. Livestock Customer Level

Customer ID	RFM Score	Level	Cluster	Color
1	5	Medium	1	Red
2	12	Low	0	Green
3	3	Medium	2	Red
10	8	Medium	1	Red
13	7	High	2	Blue

Table 5 examines the relationship between monetary value and purchase frequency across various product segments. The data demonstrates that customers who spend more money generally do so less frequently, whereas those with lower spending tend to have higher purchase frequencies. This trend suggests that high-value purchases might be associated with a preference for bulk buying or premium products. Conversely, frequent low-value transactions might indicate budget-conscious behavior or preference for smaller, more frequent purchases. These insights can guide the development of tiered marketing strategies, such as exclusive offers for high spenders and frequent buyer discounts for low monetary customers, tailored to enhance customer retention and boost revenue.

Handicraft Products

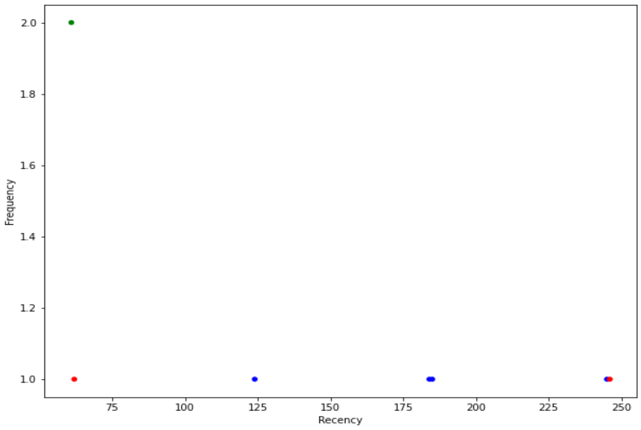

Figure 12. Handicraft Products Spread Diagram

Figure 12 shows a scatter plot analyzing the relationship between recency and frequency for customers in the handicraft product category. The scatter plot reveals that customers with lower recency values tend to have higher transaction frequencies, suggesting that these customers are actively engaged and regularly purchase handicrafts. In contrast, customers with higher recency values generally exhibit lower frequencies, indicating less frequent engagement. The spread of data points also indicates a diverse range of customer behaviors, from highly frequent buyers who make regular purchases to those who might be occasional buyers. This plot is crucial for identifying key customer segments that require different marketing approaches: frequent buyers may be interested in loyalty programs or exclusive offers, while less frequent buyers may need incentives or reminders to encourage more regular purchasing. Analyzing these patterns enables businesses to develop effective marketing strategies that address the specific needs and behaviors of their customer base, ultimately driving higher engagement and sales in the handicraft product segment.

Table 6. Handicraft Products Customer Level

Customer ID	RFM Score	Level	Cluster	Color
1	5	Medium	0	Red
2	12	Low	0	Green
10	3	Medium	0	Red
18	8	Medium	2	Red
23	7	High	2	Blue

Table 6 illustrates the optimal number of clusters (K) derived from the K-Means clustering algorithm using square distances. The table shows that the optimal K value is three, indicating that customers can be effectively grouped into three distinct clusters based on their RFM scores. This clustering approach allows for more precise customer segmentation, facilitating targeted marketing strategies for each group. The identified clusters—low, medium, and high—represent different levels of engagement and spending behavior. For example, low-level customers might need more incentives to increase their purchases, medium-level customers could be encouraged to explore new product categories, and high-level customers might benefit from loyalty programs to reinforce their purchasing behavior. This clustering strategy provides a robust foundation for optimizing marketing efforts and enhancing overall customer engagement.

CONCLUSION

Customers with a Low level are customers with low purchase order activity, but the activity in tracking products is very high in the long term, compared to the Medium level. These customers are the group with the most purchasing activity, but are consistent on only a few products in the same period or period. At the High level, this group of customers is the group that makes the most product purchases with great effort or the costs they spend to buy a product are abundant.

Marketing strategies that can be done after knowing the level of customers are such as for groups with low levels. We can provide discounts on products that they often look for, or we can evaluate our products so that they can look more attractive such as explaining the advantages of the product and so on. At the medium level by knowing their activity they place a lot of orders but only on certain products, we can combine products for one purchase, of course at the appropriate price, the result is products that they previously rarely sold, can often be sold by combining purchase products. Then at the high level we can give rewards milestones such as member cards that they can use to get the appropriate discount.

In the future, we hope that this research method can be applied directly to the e-commerce application that has been created, so that sellers can later monitor the prediction of their sales results

Incorporating this research method directly into our e-commerce application can significantly benefit both sellers and customers. By automatically segmenting users based on their real-time purchase behavior, the application can recommend relevant products and suggest targeted marketing campaigns, ultimately enhancing user experience and driving sales. Additionally, this approach can empower sellers to develop customer retention strategies by identifying and engaging high-value customers through loyalty programs.

REFERENCES

[1] C. Kleisiari, M.-N. Duquenne, and G. Vlontzos, "E-Commerce in the Retail Chain Store Market: An Alternative or a Main Trend?," *Sustainability*, vol. 13, no. 8, p. 4392, Apr. 2021, doi: 10.3390/su13084392.

[2] I. O. Pappas, "User experience in personalized online shopping: a fuzzy-set analysis," *EJM*, vol. 52, no. 7/8, pp. 1679–1703, Jun. 2018, doi: 10.1108/EJM-10-2017-0707.

[3] N. H. Bur, I. Asfah, and A. Wahid, "Analysis of The Application of Market Digitalization at Kampung Baru Public Market Makassar City," 2023.

[4] V. Babenko, Z. Kulczyk, I. Perevosova, O. Syniavska, and O. Davydova, "Factors of the development of international e-commerce under the conditions of globalization," *SHS Web Conf.*, vol. 65, p. 04016, 2019, doi: 10.1051/shsconf/20196504016.

[5] D. Voramontri and L. Klieb, "Impact of social media on consumer behaviour," *IJIDS*, vol. 11, no. 3, p. 209, 2019, doi: 10.1504/IJIDS.2019.101994.

[6] M. Molinaro, P. Romano, M. Battistutta, and G. D. Re, "An interactive DSS to improve decision-making in the (r,Q) policy," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 1438–1443, 2019, doi: 10.1016/j.ifacol.2019.11.401.

[7] M. Yahya, J. M. Parenreng, F. Fathahillah, M. S. N. Wahid, M. Fajar B, and A. Wahid, "A Decision Support System to Determine the Familiy's Economic Status for Certificate of The Low-Income Household Using MAUT Method," *ISI*, vol. 7, no. 2, p. 185, Nov. 2022, doi: 10.35314/isi.v7i2.2439.

[8] Z. Yang, Y. Shi, B. Wang, and H. Yan, "Website Quality and Profitability Evaluation in Ecommerce Firms Using Two-stage DEA Model," *Procedia Computer Science*, vol. 30, pp. 4–13, 2014, doi: 10.1016/j.procs.2014.05.375.

[9] J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," *J. Phys.: Conf. Ser.*, vol. 1142, p. 012012, Nov. 2018, doi: 10.1088/1742-6596/1142/1/012012.

[10] E. Ernawati, S. S. K. Baharin, and F. Kasmin, "A review of data mining methods in RFM-based customer segmentation," *J. Phys.: Conf. Ser.*, vol. 1869, no. 1, p. 012085, Apr. 2021, doi: 10.1088/1742-6596/1869/1/012085.

[11] M. Sarkar, A. R. Puja, and F. R. Chowdhury, "Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis," *JBMS*, vol. 6, no. 2, pp. 54–60, Mar. 2024, doi: 10.32996/jbms.2024.6.2.5.

[12] M. T. Ballestar, P. Grau-Carles, and J. Sainz, "Customer segmentation in e-commerce: Applications to the cashback business model," *Journal of Business Research*, vol. 88, pp. 407–414, Jul. 2018, doi: 10.1016/j.jbusres.2017.11.047.

[13] M. F. Faraone, M. Gorgoglione, C. Palmisano, and U. Panniello, "Using context to improve the effectiveness of segmentation and targeting in e-commerce," *Expert Systems with Applications*, vol. 39, no. 9, pp. 8439–8451, Jul. 2012, doi: 10.1016/j.eswa.2012.01.174.

[14] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *JNKTI*, vol. 3, no. 3, pp. 187–201, Dec. 2020, doi: 10.32672/jnkti.v3i3.2428.

[15] J. A. Fehrer, H. Woratschek, C. C. Germelmann, and R. J. Brodie, "Dynamics and drivers of customer engagement: within the dyad and beyond," *JOSM*, vol. 29, no. 3, pp. 443–467, Jun. 2018, doi: 10.1108/JOSM-08-2016-0236.

[16] S. Stremersch and G. J. Tellis, "Strategic Bundling of Products and Prices: A New Synthesis for Marketing," *Journal of Marketing*, vol. 66, no. 1, pp. 55–72, Jan. 2002, doi: 10.1509/jmkg.66.1.55.18455.

[17] A. Chandramouli, "Strategies and Impact of Customer Loyalty Programs in the Technology Sector: A Comprehensive Analysis," *Open Access*, vol. 1, no. 2, 2020.

[18] M. Nabipour, P. Nayyeri, H. Jabani, S. S., and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020, doi: 10.1109/ACCESS.2020.3015966.

[19] H. Hu, L. Tang, S. Zhang, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with Google Trends," *Neurocomputing*, vol. 285, pp. 188–195, Apr. 2018, doi: 10.1016/j.neucom.2018.01.038.

[20] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019, doi: 10.1073/pnas.1900654116.

[21] D. Lien Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon, "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network," *IEEE Access*, vol. 6, pp. 55392–55404, 2018, doi: 10.1109/ACCESS.2018.2868970.

[22] B. J. Ali and G. Anwar, "Marketing Strategy: Pricing strategies and its influence on consumer purchasing decision," *IJREH*, vol. 5, no. 2, pp. 26–39, 2021, doi: 10.22161/ijreh.5.2.4.

[23] S. Kaur Chatrath, G. S. Batra, and Y. Chaba, "Handling consumer vulnerability in e-commerce product images using machine learning," *Heliyon*, vol. 8, no. 9, p. e10743, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10743.

[24] C. A. Suwandi, R. Yanto, and D. Apriadi, "Implementasi Metode Apriori pada Data Mining untuk Pola Pembelian Barang pada Toko Matahari Kota Lubuklinggau," vol. 03, no. 01, 2021.

[25] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI*, vol. 2, no. 1, pp. 23–32, Apr. 2018, doi: 10.21460/jutei.2018.21.76.

[26] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis Based on k-Means," *IJEECS*, vol. 18, no. 1, p. 470, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp470-477.

[27] R. Kohavi, L. Mason, R. Parekh, and Z. Zheng, "Lessons and Challenges from Mining Retail E-Commerce Data," *Machine Learning*, vol. 57, no. 1/2, pp. 83–113, Oct. 2004, doi: 10.1023/B:MACH.0000035473.11134.83.

[28] E. Ernawati, S. S. K. Baharin, and F. Kasmin, "A review of data mining methods in RFM-based customer segmentation," *J. Phys.: Conf. Ser.*, vol. 1869, no. 1, p. 012085, Apr. 2021, doi: 10.1088/1742-6596/1869/1/012085.