Journal of Information Systems and Informatics
Vol. 5, No. 1, March 2023 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v5i1.421 Published By DRPM-UBD

Customer Loyalty Analysis Using Recency,Frequency, Monetary (RFM) and K-means Cluster for Labuan Bajo Souvenirs in Online Store

Supangat[1], Yosevina Mulyani [2]

[1.2] Program Studi Informatika, Universitas 17 Agustus 1945 Surabaya, Surabaya, Indonesia
Email: [1] Supangat@sby.untag-sby.ac.id, [2] 1461900040@srel.untag-sby.ac.id

Abstract

An average Labuan Bajo souvenir shop sells a variety of souvenirs specific to Labuan Bajo. However, the sales process is still manual, with the shop resorting to telephone or WhatsApp communication to connect with customers for placing orders. To increase sales, typical souvenir shops in Labuan Bajo are interested in adopting effective marketing strategies. Consequently, an automated system is necessary to manage customers. The Recency, Frequency, and Monetary Analysis methods are commonly used for assigning values or weights to customers during transactions. These weights are then analyzed and grouped using k-means. Recent data analysis over the last three months reveals that the typical Labuan Bajo souvenir shop has one regular customer, three potential customers, and six regular customers. Testing the system's features showed that it was functioning correctly, and therefore, it can assist the typical Labuan Bajo souvenir shop in streamlining the sales process.

Keywords: online shop, RFM, K-Means, customer segmentation.

1. INTRODUCTION

A typical Labuan Bajo souvenir shop offers a range of souvenirs specific to Labuan Bajo. Unfortunately, the gift shop faces a message stacking problem via WhatsApp, resulting in numerous errors during order recording. Furthermore, the increasing competition in selling typical Labuan Bajo souvenirs has prompted the store to devise effective marketing strategies. The store aims to enhance customer service by introducing special discount programs based on transactions to boost sales and capture customer attention [1]. To address these issues, the store can implement a computer system connecting customers and sellers to simplify ordering and data recording. The system can also analyze each customer's order history using the recency, frequency, and monetary (RFM) method. This analysis will generate a dataset that the K-means algorithm will

cluster. The results will classify customers into predetermined groups and offer special discounts to attract and retain customers, increasing orders.

RFM is a widely used method that categorizes data based on purchase novelty (recency), purchase frequency (frequency), and total purchase price (monetary) [2]. It helps to observe various customer behaviors [3]. Recency refers to the time span between the last purchase transaction and the data retrieval date, be it daily, monthly, or yearly. Frequency is the frequency with which customers make purchase transactions. Monetary is the amount spent by customers to make purchase transactions [4].

Previous studies have utilized the K-Means and RFM clustering methods for various purposes, such as determining potential and loyal customers through customer segmentation using the RFM model and K-Means to identify suitable attributes for customer segmentation [5]. In addition, other studies have been conducted to determine customer groups with high toy and profitability values that are potentially profitable for the company. The segmentation process begins with data analysis, which is then transformed using LRFM (length, recency, frequency, and monetary) and classified using the fuzzy C-Means method [6].

One case study conducted at PT Coversuper Indonesia Global utilized the RFM method in determining market segmentation by identifying the characteristics of each individual, allowing businesses to better understand their customers and create targeted marketing strategies [7]. Similarly, the RFM method has been used to help store business managers determine journal entries for clothing product restock requirements based on the transaction date, color, size, and total revenue [8].

The aim of this study is to conduct an in-depth analysis of customer loyalty for Labuan Bajo souvenirs in an online store using the Recency, Frequency, Monetary (RFM), and K-means Cluster methods. By utilizing these methods, this study seeks to examine the purchasing behavior of customers and their loyalty towards Labuan Bajo souvenirs in the online store. The results of this analysis will provide valuable insights into how to improve customer loyalty, increase sales, and attract new customers to the store. The findings of this study may also serve as a reference for future studies on customer loyalty analysis using RFM and K-means Cluster methods for other types of online stores.

2. RESEARCH METHODS

This study consists of five stages: problem identification, data collection, data processing (selection, processing, and clustering), design and code implementation, and testing, as shown in Figure 1.

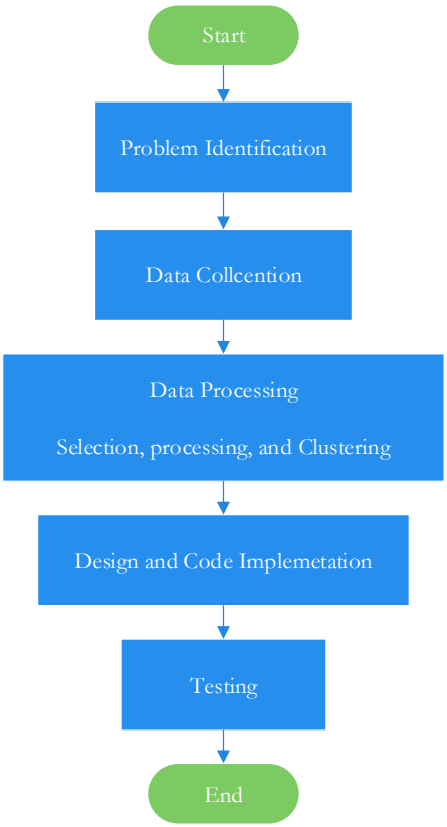

Figure 1 Research Stage

The thinking framework consists of five stages, which are described as follows:

1. Problem identification stage: At this stage, a process is conducted to identify the problems that occur at the research location. This information will later be used in the customer loyalty analysis research process using the recency, frequency, and monetary (RFM) method and K-means clustering at a typical gift shop in Labuan Bajo.
2. Data collection stage: The data collection process was carried out at this stage by interviewing the owner of the Labuan Bajo souvenir shop regarding requests for permits and data collection in the research process.
3. Data processing stage: The data processing process is carried out at this stage based on the obtained data. The data management stage begins with the following processes:

a. Selection: At this stage, the initial dataset was taken from sales data. The sales data was obtained from interviews with typical Labuan Bajo souvenir shop owners. The necessary data for this research was collected from several existing sales tables and combined.
b. Preprocessing: This stage is used to prepare the dataset so that it has better quality and is more effective before modeling. At this stage, feature selection is carried out using the RFM method (feature selection by RFM), resulting in three attributes: recency, frequency, and monetary. The data description stage is continued with the equal width technique, which gives weight to each continuous data feature to make it discrete. Thus, a customer dataset with simple features is obtained with the following attributes: customer ID, recency, frequency, and monetary.
c. Clustering: The customer dataset clustering uses KMeans with a value of K=3. The value of K = 3 is used to group customers according to the weighting carried out by the RFM technique, i.e., customer weighting, which is categorized with a value of 3, 2, and 1. The conversion of 3 = "Very Loyal," 2 = "Loyal," and 1 = "Potential Loyal.

4. Design and implementation stage: At this stage, the system design is carried out based on the problems identified. Therefore, in this study, a system was created for analyzing customer loyalty using the recency, frequency, monetary (RFM) and K-means cluster methods at Labuan Bajo souvenir shops. Based on the collected data, it is made into a web-based application.
5. Testing stage: In the final stage, BlackBox testing is carried out on the web-based application with the recency, frequency, monetary (RFM) and K-means cluster models to check if the system is suitable, the existing functions, and if it is running correctly according to the purpose of this research.

3. RESULTS AND DISCUSSION

3.1 Customer Grouping Analysis

At this stage, it is pertinent to note that the author has attached the comprehensive results of the data collection process conducted at a typical Labuan Bajo souvenir shop. This shop is located in the picturesque town of Labuan Bajo, situated in the West Manggarai Regency of the stunning East

Nusa Tenggara Province. The data has been collected for the period spanning the last three months of July through September 2022. The results of this customer transaction data collection have been meticulously sorted and organized into a comprehensive dataset based on the recency, frequency, and monetary (RFM) model. The dataset will serve as a foundation for the subsequent analysis of customer loyalty using the RFM and K-means cluster analysis techniques, thus providing valuable insights for enhancing the overall customer experience of the souvenir shop. Data grouping will analysis as presented in Table 1.

Table 1 Data Collection Results

IDCus	R	F	M
Modoku	28	3	1,200,000.00
Labuan Square	9	21	8,400,000.00
Bajo Bakery	17	14	6,300,000.00
Theresa	1	4	3,000,000.00
Surya Agung	14	27	10,500,000.00
Kado Bajo	22	9	5,200,000.00
Denis Mart	21	12	6,820,000.00
Exotic Komod	7	32	13,120,000.00
Ayana Komod Resort	12	18	7,000,000.00
Artomoro	32	7	8,650,000.00

3.2 Customer Data Management

At this stage, an analysis of data processing is being conducted, starting with the identification of the dataset based on the transaction criteria pattern using the Recency, Frequency, and Monetary (RFM) models.

1. Recency Value Calculation, Frequency, Monetary (RFM)

Guided by the Recency, Frequency, Monetary (RFM) model which consists of three criteria, namely Recency, Frequency, Monetary, data extraction is carried out with the attributes needed to measure the value of customer loyalty. The data used is data that has passed the collection stage. and data processing which is then carried out by the clustering process [9].

a. Recency = Transaction reference date
b. Frequency = Number of transactions
c. Monetary = Value of money transactions

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2. Transaction Grouping Pattern Data

The pattern of criteria used to classify customer transactions is Recency, Frequency, and Monetary. In this preprocessing process, the discretization stage is carried out, namely the process of giving weight to each continuous data feature so that it becomes categorical discrete data. [10] Thus a customer dataset with simple features is obtained. RFM pattern analysis is influenced by the following factors:

The weight value of each pattern determined by the Labuan Bajo Typical Gift Shop is yes by giving a weight value between the time range 1-5 with the conversion 5 = "Very Satisfied", 4 = "Satisfied", 3 = "Ordinary", 2 = "Not Satisfied", 1=" Not Satisfied". The part with the largest ratio value is given a weight of 5 because it indicates that the customer is increasingly satisfied. And so on until the customer with the smallest ratio value is given a weight of 1. The results of this weighting become the basic dataset to be processed for further processing, namely customer grouping using K-Means. The following is an example of data that has gone through the collection process.

Table 2 Weighting Table

No	Pattern	Sub Pattern	Scor
1	Recency	Make a Repurchase 1- 7 Days	5
		Make a Repurchase 8-14 Days	4
		Make a Repurchase 15 – 21 Days	3
		Make a Repurchase 22 – 28 Days	2
		Make a Repurchase 29-35 Days	1
2	Frequency	Number of Orders 1- 8 Transactions / Month	1
		Number of Orders 9 -14 Transactions / Month	2
		Number of Orders 15 – 20 Transactions/ Month	3
		Number of Orders 21 – 26 Transactions / Month	4
		Number of Orders 29-32 Transactions / Month	5
3	Monetary	Order Value <3,583,000.00	1
		Order Value 3,584,000.00 – 5,967,999.00/ Month	2
		Order Value 5,968,000.00 –	3

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

No	Pattern	Sub Pattern	Scor
		8,351,999.00/Month Order Value 8,352,000.00 – 10,735,999.00/Month	4
		Order value > 10,736,000.00/month	5

From analysis of recency, frequency, and monetary (RFM) patterns. The criteria set by the typical Labuan Bajo souvenir shop are as follows:

Table 3 Customer Criteria

No	Criteria	Recency (Day)	frequency (Transactions/Month)	Monetary (RP/Month
1	Regular customer	1-7 Days	29-32	> 10,736,000.00
2	Potential Customers	8-14 Days	21–26	8,352,000.00 – 10,735,999.00
3	Customer	>15 Days	<20	< 8,351,999.00

With the criteria that have been set, the customers of the Labuan Bajo Souvenir Shop provide bonuses according to predetermined criteria to segmented customers. Here are the bonus conditions for the Labuan Bajo Souvenir Shop.

Table 4 Bonus Criteria

No	Criteria	Bonus
1	Regular customer	20% of the transaction on goods
2	Potential Customers	10% of goods transactions
3	Regular Customer	5% of the transaction on goods

The following is a Data Set of weighted transactions based on Recency,Frequency , Monetary (RFM).

Table 5 RFM Datasets

IDCus	R	F	M
Modoku	2	1	1
Labuan Square	4	4	4
Bajo Bakery	3	2	3
Theresa	5	1	1

IDCus	R	F	M
Surya Agung	4	5	4
Kado Bajo	2	2	2
Denis Mart	3	2	3
Exotic Komod	5	5	5
Ayana Komod Resort	4	3	3
Artomoro	1	1	4

3.3 Clustering Stage

Next at this stage is the data processing stage with the k-means algorithm to group customers into the same characteristics.

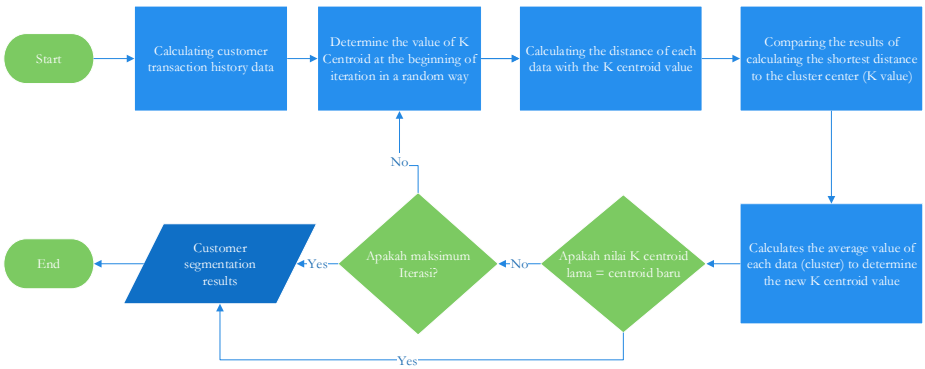

Figure 2. Clustering Stage

The steps for analysis using the K-Means method begin by collecting customer transaction history data. For example, the following is customer performance data in predetermined weights [11].

Table 6 Distributor Data

IDCus	R	F	M
Modoku	2	1	1
Labuan Square	4	4	4
Bajo Bakery	3	2	3
Theresa	5	1	1
Surya Agung	4	5	4
Kado Bajo	2	2	2
Denis Mart	3	2	3

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Exotic Komod	5	5	5
Ayana Komod Resort	4	3	3
Artomoro	1	1	4

Furthermore, in the analysis process, the K-Means method is used to determine the K value for the centroid or cluster center. The cluster center in this study is determined by finding the average value of the criteria weight from the customer performance data. The K value for each criterion for iterations is shown below.

Table 7 Early Centroids

Centroids 1	4	3	3
Centroids 2	4	4	4
Centroids 3	5	5	5

Then perform the distance calculation technique using the Eucledian Distance with the formula in Aquetion 1.

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{n} (x_{il} - x_{jl})^2} \quad (1)$$

Table 8 Compute Euclidean Distance Iteration 1

IDCus	R	F	M	C1	C2	C3
Modoku	2	1	1	10	20	35
Labuan Square	4	4	4	2	0	3
Bajo Bakery	3	2	3	2	6	15
Theresa	5	1	1	9	19	32
Surya Agung	4	5	4	5	1	2
Kado Bajo	2	2	2	4	10	21
Denis Mart	3	2	3	2	6	15
Exotic Komod	5	5	5	9	3	0
Ayana Komod Resort	4	3	3	0	2	9
Artomoro	1	1	4	8	12	21

Once the Euclidean distance is known, the C1, C2, and C3 values can be identified, namely the closest distance to each data object with cluster values of

1, 2, and 3. Then determine the shortest distance from an object to the centroid point using the minimum value formula between C1 and C2 and C3. Calculations can be continued to the 2nd iteration process by determining the value of the new centroid point by calculating the average for each cluster using the following equation.

$$c_{kl} = \frac{x_{1l} + x_{2l} + \ldots + x_{pl}}{p} \quad (2)$$

Table 9 New Cluster

Centroids 1	2.857143	1.714286	2.428571
Centroids 2	4	4,5	4
Centroids 3	5	5	5

From the results of the new centroid, then calculate the shortest distance from each data object to the centroid using the Euclidean Distance formula.

Table 10 Compute Euclidean Distance 2nd Iteration

IDCus	R	F	M	**C1**	**C2**	**C3**
Modoku	2	1	1	1.812654	5.024938	6.403124
Labuan Square	4	4	4	3	0.5	1.732051
Bajo Bakery	3	2	3	0.654654	2.872281	4.123106
Theresa	5	1	1	2.672612	4.716991	5.656854
Surya Agung	4	5	4	3.817254	0.5	1.414214
Kado Bajo	2	2	2	1	3.774917	5.196152
Denis Mart	3	2	3	0.654654	2.872281	4.123106
Exotic Komod	5	5	5	4.690416	1.5	0
Ayana Komod Resort	4	3	3	1.812654	1.802776	3
Artomoro	1	1	4	2.535463	4.609772	5.744563

Once the Eucledian Distance is known, the C1, C2, C3 values can be identified, namely the closest distance to each data object with cluster values of 1, 2, 3. If the cluster results in iteration 1 are different from the clusters in iteration 2, repeat the k-means process with newer centroids. To determine the new cluster center (centroid) by calculating the average of each cluster using the equation:

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

$$c_{kl} = \frac{x_{1l} + x_{2l} + \ldots + x_{pl}}{p} \quad (3)$$

Table 11 New Centroids

Centroids 1	2.666667	1.5
Centroids 2	4	4
Centroids 3	5	5

From the results of the new centroid, then calculate the shortest distance from each data object to the centroid using the Euclidean Distance formula.

Table 12 Final iteration results

IDCus	R	F	M	C1	C2	C3
Modoku	2	1	1	1.57233	4.484541	6.403124
Labuan Square	4	4	4	3.28718	0.333333	1.732051
Bajo Bakery	3	2	3	0.897527	2.333333	4.123106
Theresa	5	1	1	2.733537	4.136558	5.656854
Surya Agung	4	5	4	4.099458	1.054093	1.414214
Kado Bajo	2	2	2	0.897527	3.282953	5.196152
Denis Mart	3	2	3	0.897527	2.333333	4.123106
Exotic Komod	5	5	5	4.980518	1.943651	0
Ayana Komod Resort	4	3	3	2.114763	1.20185	3
Artomoro	1	1	4	2.409472	4.255715	5.744563

This calculation process is repeated until the iteration results are found to be the same as the previous iteration and in this condition the calculation can be stopped because the cluster conditions have reached convergence.

Table 13 Number of Labuan Bajo Typical Souvenir Shop Clusters

Membership of Each Cluster		
Clusters	**Status**	**Amount**
3	Very Loyal	1
2	Loyal	3
1	Normal	6

Cluster calculation results show that cluster 3 with "Highly Loyal" status has a total of 1, cluster 2 with "Loyal" status has a total of 3, and cluster 1 with "Ordinary" status has a total of 6. The research stage is an overview that explains the logical flow of the research in general. The following is a picture and description of the stages carried out within the research framework, including:

3.4 System Implementation

The system that has been developed, in the form of a website, will be demonstrated in the following figures. Upon opening the website, users will be directed to a login page for those who already have an account or a registration page for new users who need to create an account. Once the user logs in using their account, they will be redirected to the main page of the website, where details of the products available for purchase will be displayed, as depicted in Figure 3.

Figure 3 Page of RFM Analysis and K-means Iteration 1

The webpage provides a comprehensive display of the results derived from the RFM and K-means analysis. In Iteration 1, the findings are represented in the form of customer grouping or classes, which can be accessed via the dedicated page. Additionally, the final output of the analysis can be viewed on the Result Page (Figure 4). Furthermore, the webpage also presents the Transaction Data Page (Figure 5), which provides a detailed overview of the transactional data gathered during the analysis process.

Journal of Information Systems and Informatics
Vol. 5, No. 1, March 2023
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

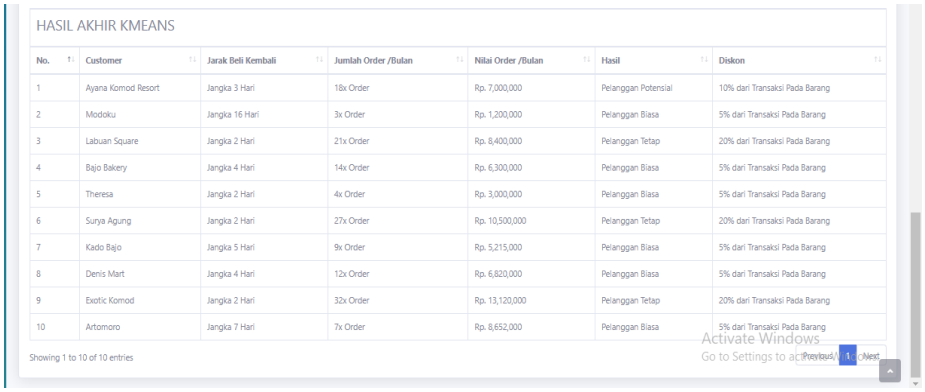

Figure 4 Final Result Page of RFM and K-means Analysis

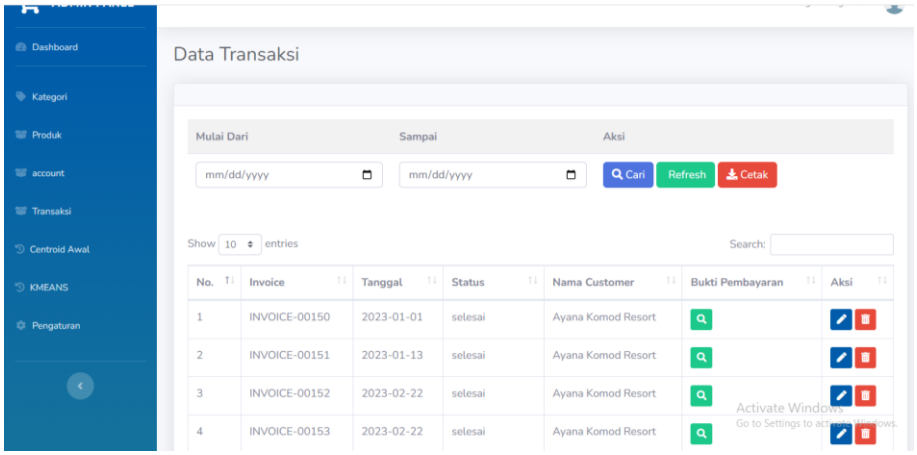

Figure 5 Transaction Data Page

3.5 Test result

The next stage of implementation is testing using the Black Box test. Black box testing is a type of software testing in which the tester examines the software application's external functionalities without having any knowledge of its internal structure, design, or coding. The tester focuses on inputs and outputs of the system and tries to identify any defects, errors, or unexpected results. This type of testing is usually performed at the end of the software development lifecycle to ensure that the software meets its functional requirements and user expectations [12]. The main objective of black box testing is to validate the software from the end-user's perspective and ensure that it behaves as expected. The black box test results are in Table 5.

Table 5. Black Box Test Result Data

Test Page	Test Category	Results
Category Data Management	Adding item category data	Succeed
	Changing item category data from the list	Succeed
	Removes item category data from the list	Succeed
Product Data Management	Adding product data to the system	Succeed
	Changing product data on the list	Succeed
	Delete product data on the list	Succeed
Customer Data Management	Enter customer additions to the system	Succeed
	Changing customer data on the list	Succeed
	Deleting customer data on the system	Succeed
Purchase	Edit customer transaction data	Succeed
	Delete transaction data on the list	Succeed
Centroid Management of K-Means	Editing the initial K-means centroid data	Succeed
Customer Segmentation Analysis	Open the K-Means analysis menu	Succeed
	Displays customer segmentation results with detailed K-Means calculations	Succeed

4. CONCLUSION

Several conclusions will be drawn from the result. Firstly, recency, frequency, and monetary (RFM) can be effectively applied to group customers in typical Labuan Bajo souvenir shops. This system is proficient in analyzing purchase loyalty, number of product purchases, and customer spending, thus assisting souvenir shops in Labuan Bajo during transactions between stores and customers. Secondly, the K-means clustering method can be employed to group customer transaction data in typical Labuan Bajo souvenir shops. The study showed that over the last 3 months, the typical souvenir shop had one regular customer, three potential customers, and six regular customers. Thirdly, the customer loyalty analysis system employs the K-means cluster algorithm to classify incoming customer transaction data and categorize it into regular, loyal, and irregular customer data sets. This system can assist typical souvenir shops in Labuan Bajo in creating effective strategies to retain regular and loyal customers while attracting new ones.

REFERENCE

[1] S. S. Prasetyo, M. Mustafid, and A. R. Hakim, "Penerapan Fuzzy C-Means Kluster Untuk Segmentasi Pelanggan E-Commerce Dengan

Metode Recency Frequency Monetary (RFM)," *Jurnal Gaussian*, vol. 9, no. 4, pp. 421–433, Dec. 2020, doi: 10.14710/j.gauss.v9i4.29445.

[2] J. Jamal and D. Yanto, "Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer," *Jurnal Energy*, vol. 9, no. 1, pp. 1–8, May 2019.

[3] W. A. Taqwin, N. Y. Setiawan, and F. A. Bachtiar, "Analisis Segmentasi Pelanggan dengan RFM Model pada PT. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 2, pp. 1986–1993, Feb. 2019.

[4] Sudriyanto, "Clustering Loyalitas Pelanggan dengan Metode RFM (Recency, Frequency, Monetary) dan Fuzzy C-Means," in *SNATIF*, 2017, pp. 815–822.

[5] Sandi, T. A. Armawan, M. Raharjo, J. L. Putra, and R. Ridwan, "Clustering Kesetian Pelanggan denan Model RFM (Recency, Frequency, Monetary) dan K-Means," *Jurnal Pilar Nusa Mandiri*, vol. 14, no. 2, 2018.

[6] Waroka, Lia, S. Monalisa, D. Anjainah, and N. Arifin, "Implementasi Algoritma Fuzzy C-Means (FCM) dalam Pengklasterisasian Nilai Hidup Pelanggan Dengan Model LRFM," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 6, no. 1, 2020.

[7] A. T. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *Konstelasi: Konvergensi Teknologi dan Sistem Informasi*, vol. 1, no. 1, pp. 204–215, Apr. 2021, doi: 10.24002/konstelasi.v1i1.4293.

[8] R. Y. Firmansah, J. Dedy Irawan, and N. Vendyansyah, "Analisis Rfm (Recency, Frequency And Monetary) Produk Menggunakan Metode K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 5, no. 1, pp. 334–341, Feb. 2021, doi: 10.36040/jati.v5i1.3282.

[9] Novendri, "Pengertian Web," *Lentera Dumai*, vol. 10, no. 2, pp. 46–57, 2019.

[10] Supriyono, "Software Testing with the Approaach of Blackbox Testing on the Academic Information System," *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 227–233, 2020.

[11] N. N. N. Nurliza, "Penerapan Euclidean Distance pada Pengenalan Pola Citra Sidik Jari," Universitas Islam Indonesia, Yogyakarta, 2018.

[12] S. Ahmad, and I. Effendy. "Implementasi Model View Controller (Mvc) Pada Pengembangan Sistem Informasi Pengaduan Masyarakat Pada Polres Pagaralam." *Jurnal Informatika* 20, no. 2 (2020): 199-208. doi: 10.30873/ji.v20i2.2393