ILKOM Jurnal Ilmiah Vol. 15, No. 1, April 2023, pp. 32-44
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Implementation of RFM Method and K-Means Algorithm for Customer Segmentation in E-Commerce with Streamlit

Farrikh Alzami [a,1,*]; Fikri Diva Sambasri [a,2]; Mira Nabila [a,3]; Rama Aria Megantara [a,4]; Ahmad Akrom [a,5]; Ricardus Anggi Pramunendar [a,6]; Dwi Puji Prabowo [a,7]; Puri Sulistiyawati [a,8]
[a]*Universitas Dian Nuswantoro, jl. Imam Bonjol 207, Semarang, 50131, Indonesia*
[1]*alzami@dsn.dinus.ac.id;* [2]*fikri.sambasri@gmail.com;* [3]*111201911805@mhs.dinus.ac.id;* [4]*aria@dsn.dinus.ac.id;*
[5]*ahmad.akrom@dsn.dinus.ac.id;* [6] *ricardus.anggi@dsn.dinus.ac.id;* [7]*dwi.puji.prabowo@dsn.dinus.ac.id;* [8] *puri.sulistiya@dsn.dinus.ac.id*
** Corresponding author*

Article history: Received November 29, 2022; Revised December 09, 2022; Accepted January 24, 2023; Available online April 07, 2023

Abstract

E-commerce is selling and buying goods through an online or online system. One of the business models in which consumers sell products to other consumers is the Customer to Customer (C2C) business model. One thing that needs to be considered in the business model is knowing the level of customer loyalty. By knowing the level of customer loyalty, the company can provide several different treatments to its customers to maintain good relationships with customers and increase product purchase revenue. In this study, the author wants to segment customers on data in E-commerce companies in Brazil using the K-Means clustering algorithm using the RFM (Recency, Frequency, Monetary) feature and display it in the form of a dashboard using the Streamlit framework. Several stages of research must be carried out. Firstly, taking data from the open public data site (Kaggle), then merging the data to select some data that needs to be used, understanding data by displaying it in graphic form, and conducting data selection to select features/attributes. The step follows the proposed method, performs data preprocessing, creates a model to get the cluster, and finally displays it as a dashboard using Streamlit. Based on the results of the research that has been done, the number of clusters is 4 clusters with the evaluation value of the model using the silhouette score is 0.470.

Keywords: E-Commerce; Customer Segmentation; K-Means; RFM; Streamlit.

Introduction

Based on a survey by the Association of Internet Service Providers in 2018, 64,8% of internet usage has increased and used in various industrial or individual sectors[1]. It is recorded that the island of Java in Indonesia has a contribution of large internet users 55% [2]. The use of internet technology has several advantages, including the ability to sell or buy items, find information about the present situation, learn tools to expand knowledge, and communicate with people. The internet benefits a variety of industries, including the electronic commerce sector (E-commerce)[3], [4].

E-commerce is the practice of purchasing and reselling products online. Customer-to-customer (C2C) is one of the e-commerce business models in which customers of an online marketplace sell goods to other customers. [5]. One example of an e-commerce business model in this research is Olist e-commerce (https://olist.com/). Olist is Brazil's largest department store (marketplace), which operates in the e-commerce segment and is founded in 2015 [6]. Olist has a total sale of 12 million with 6,5 million products registered. Determining the extent of consumer loyalty relates to the e-commerce company strategy. Knowing the level of client loyalty allows businesses to approach customers differently, preserving good customer relations and boosting product sales. [7].

RFM segmentation is one of the techniques used to assess the extent of client loyalty. (Recency, Frequency, Monetary) [8], [9]. RFM is one way for consumer segmentation based on transaction history [10],[11] using the three criteria of recency, frequency, and monetary value. The recency parameters are the difference between the day the data analysis is to be done and the day the customer last made a transaction. This research used a day unit. The frequency parameter is the number of transactions made by customers. The monetary parameter is the total number

http://dx.doi.org/10.33096/ilkom.v15i1.1524.32-44

of orders taken out by customers. The algorithm often used for grouping (segmentation) consumers is clustering. Several algorithms are used for clustering, such as the KMeans method[12], [13], the Agglomerative clustering method, and the DBSCAN method. The K-Means method is a data mining clustering method that manages to group the data into one or more groups [14].

Previous research by Rahma Wati Br Sembiring Berahmana, Fahd Agodzo Mohammed, and Kankamol Chairuang using the K-Means algorithm and the RFM method for customer segmentation has revealed that the K-Means method has a high level of validity, with the results of the Davies Bouldin Index being 0,33009058. The Silhouette Index is 0,912671056, and the number of the best cluster is 2 [15]. Thus, this research wants to know the customer segmentation uses the K-Means clustering method by considering feature construction into RFM on the E-Commerce Olist dataset. Another hope is that UMKM must be empowered with competitiveness; one of them must understand the customer because UMKM does not have a system, then the Olist case study was chosen, which is a marketplace in Brazil. Another reason the author needs help getting the marketplace data in Indonesia, such as Tokopedia and others. The results of this research will be used as a reference for developing the UMKM RFM system (Recency, Frequency, Monetary) in Central Java.

Method

This research uses the RFM method (Recency, Frequency, and Monetary) to create features on the dataset and the K-Means clustering algorithm as a model for creating clusters to do customers segmentation; these are some of the stages carried out in this research which can be described in **Figure 1**:

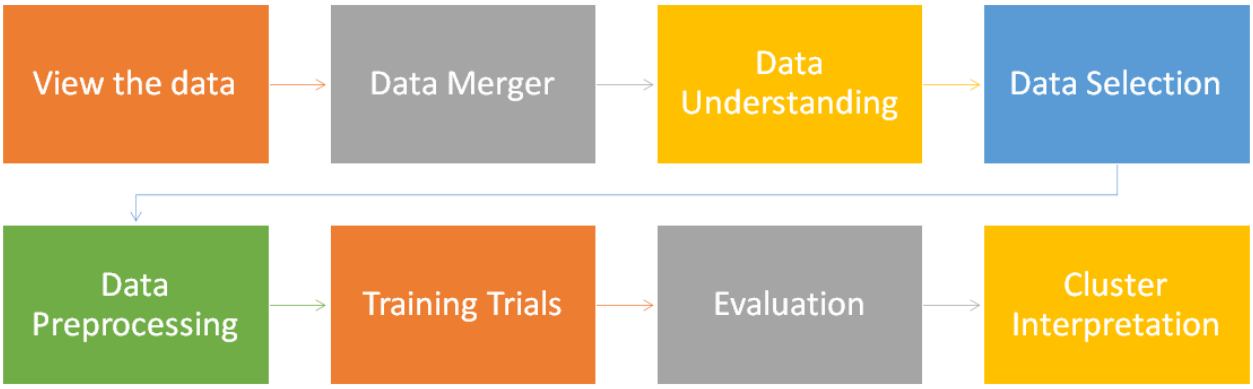

Figure 1 Process Method of Research

From **Figure 1**, we can explain as follows:

A. View Data

At this stage, the data is obtained from the public open data site, namely Kaggle. Kaggle is about the data history of customer purchases in E-Commerce companies in Brazil, that as Olist E-Commerce.

B. Data Merger

After the author saw the data through the Kaggle site, the author did merge data from several data selected to become a dataset. Data combined using the Inner Join method consists of data products, product categories, order items, orders, order payments, and customers.

C. Data Understanding

The data understanding process, the author performs exploratory data analysis (EDA) which aims to ensure that there is no problem with the existing data and find patterns/ information from the results of the visualization that created by the author.

D. Data Selection

In the data selection stage, the author selects attributes or features that match the method proposed by the author, the RFM method. Data is selected for creating RFM features: order purchase timestamp, unique customer id, order status, price, freight value, and order id.

E. Data Preprocessing

After selecting the data, it is then preprocessing the data process. It includes data cleaning to check the missing value, feature engineering to create new features or RFM features through grouping, and standardization is a process that changes the feature values so that the data value range of each feature is the same.

F. Training Trials

In the trial stage of this training, the data that has been processed is preprocessing data then a training trial is carried out from cluster 2 to cluster 9 using the K-Means clustering algorithm. At this stage, the data used is from 2016-2017, while the 2018 data is used for testing.

The model training process is carried out in model training to obtain a clustering model using the K-Means clustering algorithm and the number of clusters from the previous evaluation process. The model that has been created will be saved for use in making cluster predictions from test data. These are the steps in the K-Means algorithm process work:

1. Determine the number of clusters or the amount of K to be formed
2. Establish the center point of the cluster or centroid randomly
3. Calculate the distance between each data with centroid using Euclidean Distance

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

d = distance between x and y
x = cluster center data
y = data on attributes
i = each data
n = amount of data

4. Group of data into clusters based on the closest distance to the centroid
5. Determine the position of the new centroid using the average formula for each features
6. Repeat the step three to five until the position of cluster member is not changed
7. The model is ready

G. Evaluation

Then to get the amount of the best cluster, conducted an evaluation process with using Silhouette Score. The highest Silhouette Score is the amount of cluster selected, for the next process is the training model.

H. Cluster Interpretation

At this stage, we interpret the cluster member to obtain the information for RFM segmentation.

Results and Discussion

A. View Data

In this research, data collection was carried out from open public data sites, namely Kaggle, regarding customer purchase data in an E-Commerce company in Brazil created by Olist; the dataset contains 100 thousand information history of customer purchases from 2016 - 2018.

B. Data Merger

Merge data with the inner join method, according to the data needed to be one dataset. Obtained 115878 records where consist of 22 columns which include nine features with numeric data type and 13 features with object data type/ string. This is the result of merging the tables performed as seen at **Table 1**:

Table 1. Result of Data Merger

Field	Description
Order_id	Unique key for order
Order_item_id	The serial number identifies the number of items which belongs to the same order
Product_id	Unique key for product
Seller_id	Unique key for seller

Field	Description
Shipping_limit_date	Seller delivery deadline for handle orders to logistics partners
price	Product price
Freight_value	Product shipping cost price
Customer_id	Unique key for customer
Order_status	Product delivery status
Order_purchase_timestamp	Product purchase time
Order_estimated_delivery_date	Product delivery time
Product_category_name	Product name category
Payment_sequential	Amount of payment methods done by customer
Payment_type	Payment type used customers to buy products
Payment_installments	Amount of installments selected by the customer
Payment_value	Transaction value
Customer_unique_id	Unique key for customer
Customer_zip_code_prefix	The first 5 digits of the customer postal code
Customer_city	The name of the city customer comes from
Customer_state	The customer's area comes from
Month_order_purchase	Month of product purchase from customer
Year_order_purchase	Year of product purchase from customer

C. Data Understanding

Data understanding is a process to find out information from owned data. To find out the data information, several preparations were made visualization to get the following insights:

1. Average monthly active users and amount of new customers

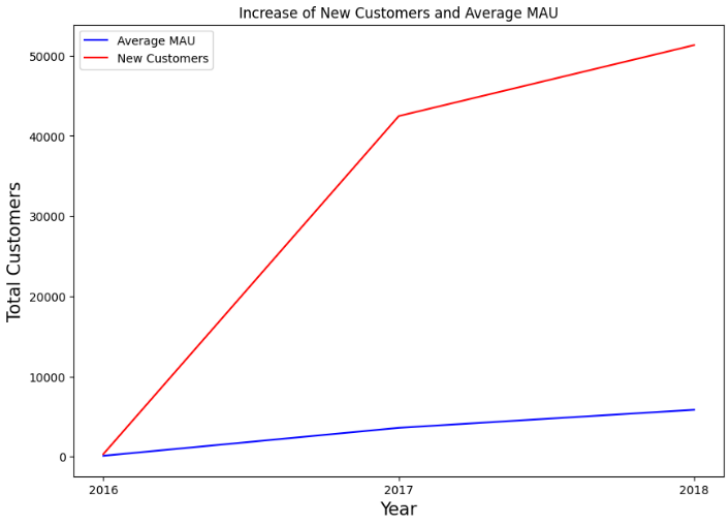

Figure 2. MAU Average & New Customers

As seen in **Figure 2**, the available data starts from transaction data in September 2016, which caused the results of the analysis in 2016 to have a significant difference compared with the value in 2017 and 2018. From this analysis, customer activities monthly and also the number of new customers have increased.

2. The number of customers making repeat orders

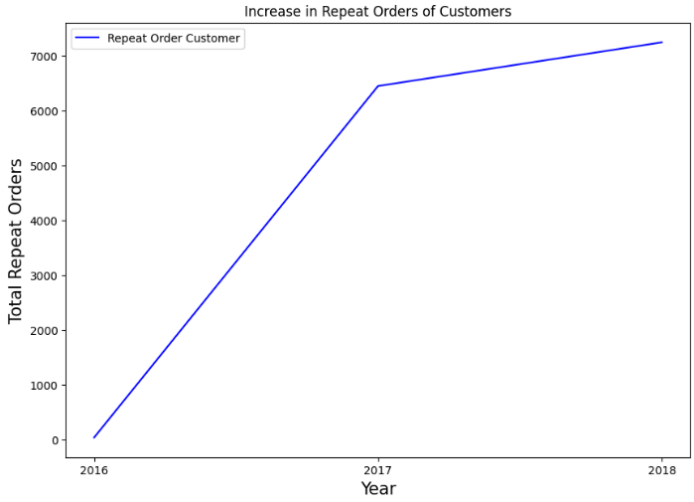

Figure 3. The Amount of Repeat Orders

In **Figure 3**, In terms of orders/ orders made by customers looks good. Seems like the number of customers who make repeat orders has increased from 2016 until 2018.

3. The amount of product category revenue per year

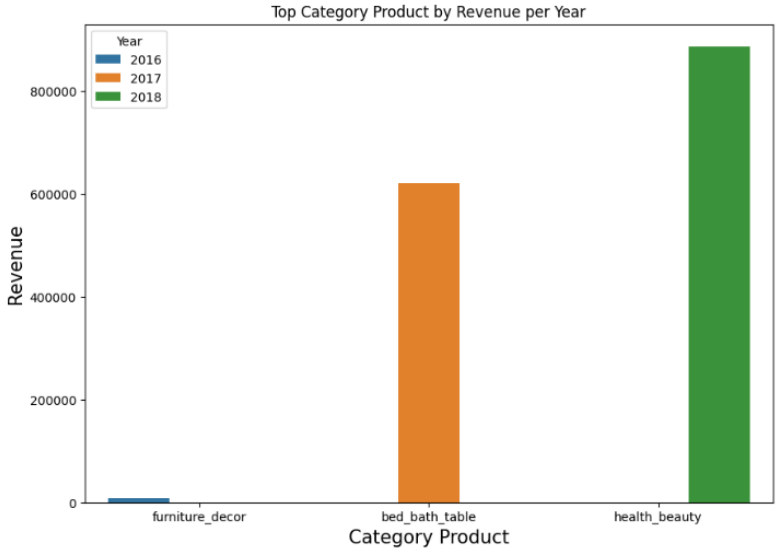

Figure 4. Product Categories Based on Highest Total Income

From the visualization at **Figure 4**, the product category that gives the highest revenue every year is always changing. Viewed from the side pf the company's overall revenue also increases every year.

4. The amount of product category cancellations per year

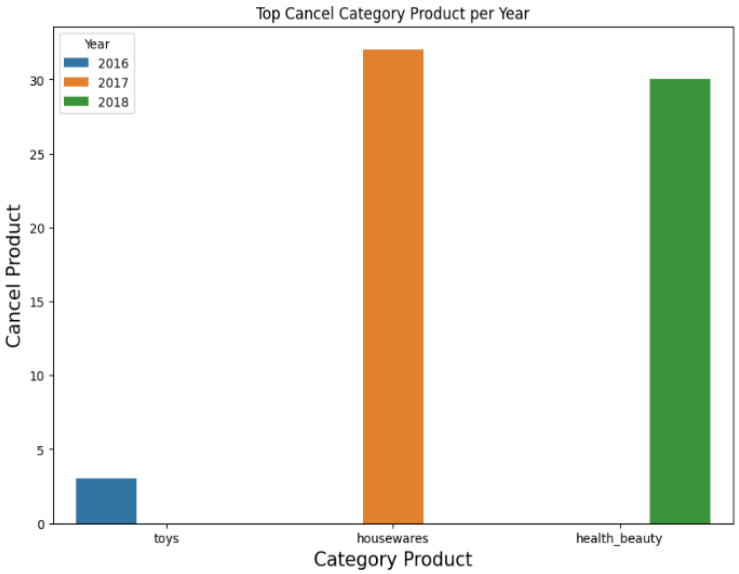

Figure 5. Product Categories Based the Highest Amount of Cancellations

From the visualization at **Figure 5**, every year the product categories have been purchased cancellations vary.

However, there is an interesting thing in 2018; the health beauty product category is the name of the product category that provides the highest revenue, as well as the name of the product category that has the highest amount of product purchase cancellations in 2018.

5. Favorite payment type per year

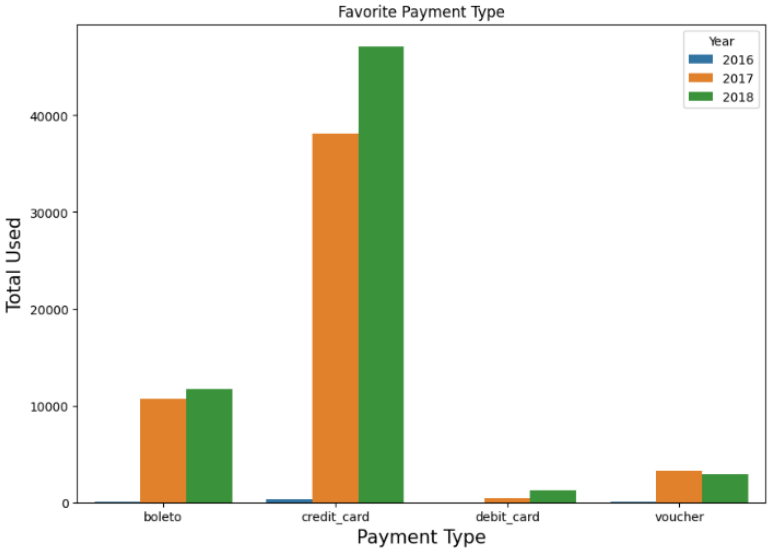

Figure 6 Favorite Payment Type

From the visualization at **Figure 6**, it is found that the preferred payment method customers in the process of purchasing products at E-Commerce companies are cards credit.

D. Data Selection

After visualizing the data to be able to understand the information in the owned data, then the appropriate feature is selected to create an RFM feature. This is example of the results of the selection features used as **Figure 7**:

	order_purchase_timestamp	customer_unique_id	order_status	price	freight_value	order_id
0	2017-09-13 08:59:02	871766c5855e863f6eccc05f988b23cb	delivered	58.9	13.29	00010242fe8c5a6d1ba2dd792cb16214
1	2017-06-28 11:52:20	0fb8e3eab2d3e79d92bb3fffbb97f188	delivered	55.9	17.96	130898c0987d1801452a8ed92a670612
2	2017-08-01 18:38:42	e7c828d22c0682c1565252deefbe334d	delivered	58.9	16.17	6f8c31653edb8c83e1a739408b5ff750
3	2017-08-10 21:48:40	0bb98ba72dcc08e95f9d8cc434e9a2cc	delivered	58.9	13.29	7d19f4ef4d04461989632411b7e588b9
4	2017-07-27 15:11:51	33449409b16400dbeaf886a5140bf59c	delivered	55.9	26.93	a0f9acf0b6294ed8561e32cde1a966bc

Figure 7. Result of Data Selections

E. Data Preprocessing

1. Data Cleaning
   Data Cleaning is the process of cleaning data from unwanted data that do not correct, incomplete, etc.
2. Feature Engineering
   Feature Engineering is the process of making existing feature into features RFM. Here is the result of creating the RFM feature can be seen at **Table 2**:

Table 2. The Result of feature Engineering

Recency	Frequency	Monetary
297	1	86.22
81	1	43.62
48	1	196.89
303	1	150.12
167	1	29.00

3. Standardization
   Standardization is the process of changing feature values that have an average value of 0 and standard deviation of 1 which can be seen at Table 3.

Table 3 The Result of Standardization

Recency	Frequency	Monetary
1.610151	-0.268343	-0.326049
-0.594647	-0.268343	-0.486358
-0.931492	-0.268343	0.090416
1.671395	-0.268343	-0.085585
0.283189	-0.268343	-0.541375

F. Modelling Process

At this stage, the researcher tries to create a model using an algorithm K-Means clustering with the number of clusters 2 to 9. Then evaluated using silhouette score to get the optimal number of clusters. Here is found the amount cluster is 4 clusters with a value of 0,470 as seen in **Figure 8**

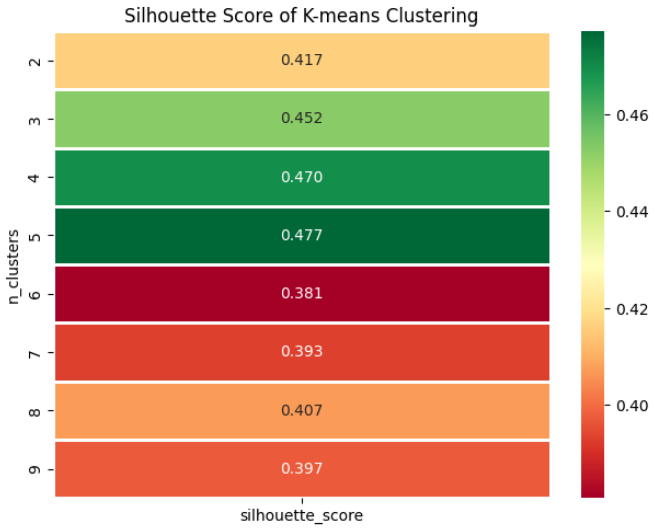

Figure 8. Silhouette Score K-Means Clustering

G. RFM Based Cluster Calculation

At this stage, the data is transformed into RFM (Recency, Frequency, Monetary), which is based on the distance between the customer's last transaction and the end of the transaction analysis period, the number of transactions made by customers during the analysis period, and the nominal amount issued by the customer. During the transformation process, standardization is to adjust the range of values for the three features.

Table 4. The Result of Data Transformation

No.	Customer Unique Id	Recency	Frequency	Monetary
1	0000f46a3911fa3c0805444483337064	1,61	-0,27	-0,33
2	0000f6ccb0745a6a4b88665a16c9f078	-0,59	-0,27	-0,49
3	0004aac84e0df4da2b147fca70cf8255	-0,93	-0,27	0,09
4	0005e1862207bf6ccc02e4228effd9a0	1,67	-0,27	-0,09
5	0006fdc98a402fceb4eb0ee528f6a8d4	0,28	-0,27	-0,54
.....				
41668	fffb09418989a0dbff854a28163e47c6	-1,27	-0,27	-0,38
41669	fffcf5a5ff07b0908bd4e2dbc735a684	0,69	0,87	7,13
41670	fffea47cd6d3cc0a88bd621562a9d061	-1,20	-0,27	-0,33
41671	ffff371b4d645b6ecea244b27531430a	1,93	-0,27	-0,23
41672	ffffd2657e2aad2907e67c3e9daecbeb	1,07	-0,27	-0,38

1. Determining the number of clusters and early centroid
   From the E-Commerce dataset, 4 groups (clusters) will be formed. Then choose the centroid randomly assigned to each cluster as seen at **Table 5**.

Table 5. Value of Early Centroid

Centroid 1				
	Data	***R***	***F***	***M***
Cluster 1	Data 1	1,61	-0,27	-0,33
Cluster 2	Data 100	-1,41	0,87	-0,15
Cluster 3	Data 205	-0,95	0,87	0,36
Cluster 4	Data 320	1,10	-0,27	0,07

2. Cluster Calculation
   Euclidean distance is a helpful way to measure the distance between points point with the midpoint (Centroid). Here is the calculation using Euclidean Distance to the dataset in 1 Centroid table with the first customer data.

$$C1(x,y) = \sqrt{(1{,}61 - 1{,}61)^2 + (-0{,}27 - (-0{,}27))^2 + (-0{,}33 - (-0{,}33))^2} = 0$$
$$C2(x,y) = \sqrt{(-1{,}41 - 1{,}61))^2 + (0{,}87 - (-0{,}27))^2 + (-0{,}15 - (-0{,}33))^2} = 3{,}23$$
$$C3(x,y) = \sqrt{(-0{,}95 - 1{,}61)^2 + (0{,}87 - (-0{,}27))^2 + (0{,}36 - (-0{,}33))^2} = 2{,}88$$
$$C4(x,y) = \sqrt{(0{,}10 - 1{,}61)^2 + (-0{,}27 - (-0{,}27))^2 + (0{,}07 - (-0{,}33))^2} = 0{,}65$$

The results of each cluster member are obtained from the value of the calculation result of Euclidean Distance smallest of each data. The following are the results of the cluster calculations obtained as seen at **Table 6**.

Table 6. The Result of Cluster Calculation Iteration 1

Iteration 1						
No	***Data***	***C1***	***C2***	***C3***	***C4***	***Results***
1	D1	0,00	3,23	2,88	0,65	1
2	D2	2,21	1,44	1,46	1,78	2
3	D3	2,58	1,25	1,16	2,03	3
4	D4	0,25	3,28	2,89	0,59	1
5	D5	1,34	2,08	1,90	1,02	4
……...						
41668	D41668	2,88	1,16	1,39	2,41	2
41669	D41669	7,60	7,57	6,97	7,16	3
41670	D41670	2,81	1,17	1,35	2,33	2
41671	D41671	0,33	3,53	3,15	0,88	1
41672	D41672	0,54	2,74	2,43	0,45	4

The results of the clustering calculations are carried out as manual calculations in iteration 1. Iteration will stop if the iteration results have a fixed clustering member. As seen at **Table 7** and **8**, on research it stops at iteration 15 as follows.

Table 7. Value of Centroid 15

Centroid 15			
	R	***F***	***M***
Cluster 1	1,43	-0,09	-0,09
Cluster 2	-0,93	-0,09	-0,10
Cluster 3	-0,005	2,54	2,96
Cluster 4	0,19	-0,10	-0,14

This is the calculation using Euclidean Distance to the dataset in the Centroid 15 table with first customer data.

$$C1(x,y) = \sqrt{(1{,}43 - 1{,}61)^2 + (-0{,}09 - (-0{,}27))^2 + (-0{,}09 - (-0{,}33))^2} = 0{,}33$$
$$C2(x,y) = \sqrt{(-0{,}93 - 1{,}61))^2 + (-0{,}09 - (-0{,}27))^2 + (-0{,}10 - (-0{,}33))^2} = 2{,}56$$
$$C3(x,y) = \sqrt{(-0{,}005 - 1{,}61)^2 + (2{,}54 - (-0{,}27))^2 + (2{,}96 - (-0{,}33))^2} = 4{,}62$$
$$C4(x,y) = \sqrt{(0{,}19 - 1{,}61)^2 + (-0{,}10 - (-0{,}27))^2 + (-0{,}14 - (-0{,}33))^2} = 1{,}44$$

The results of the calculation of the data above obtained data such as the table below.

Table 8. The Results of Cluster Iteration 15 Calculation

Iteration 15						
No	***Data***	***C1***	***C2***	***C3***	***C4***	***Results***
1	D1	0,33	2,56	4,62	1,44	1
2	D2	2,08	0,54	4,49	0,87	2
3	D3	2,38	0,26	4,12	1,16	2
4	D4	0,29	2,61	4,47	1,49	1
5	D5	1,25	1,30	4,50	0,44	4
……...						
41668	D41668	2,88	1,16	1,39	2,41	2

Iteration 15						
No	***Data***	***C1***	***C2***	***C3***	***C4***	***Results***
41669	D41669	7,60	7,57	6,97	7,16	3
41670	D41670	2,81	1,17	1,35	2,33	2
41671	D41671	0,33	3,53	3,15	0,88	1
41672	D41672	0,54	2,74	2,43	0,45	1

H. The Result of Cluster Interpretation

From the results of the modeling, it is obtained the interpretation of the cluster based on the RFM features as following:

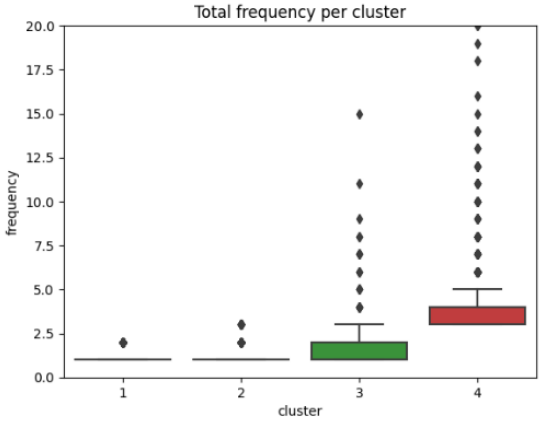

Figure 10. Frequency per Cluster

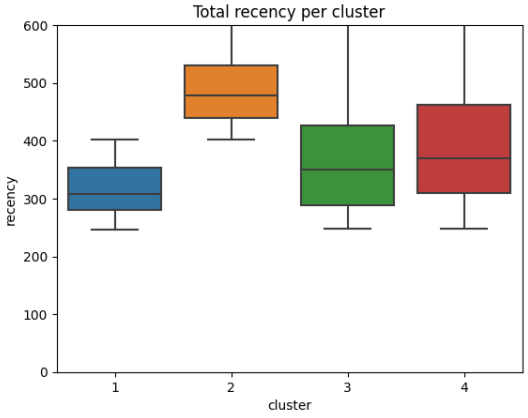

Figure 11. Recency per Cluster

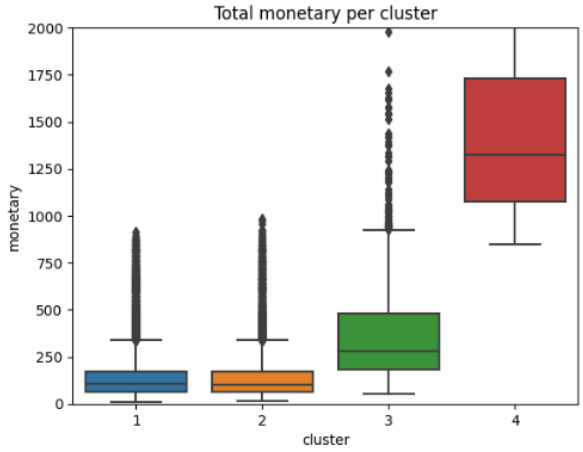

Figure 12. Monetary per Cluster

From the cluster distribution per RFM feature (Recency, Frequency, Monetary) as seen at **Figure 9**, **Figure 10**, **Figure 11**, it is known that:

1. Need Attention Customers have characteristics, namely the intensity of purchases with an average of one purchase, have an average spend of $142, and an average distance purchase history less than one year.
2. At Risk Customer has the characteristics of purchasing intensity with average one-time purchase, has an average spend of $142, and an average historical distance purchase more than one year three months.
3. Loyal customers have the characteristics of purchasing intensity with an average four purchases, $389 average spend, and average historical distance one year purchase.
4. Potential Loyalty Customers have characteristics, namely purchase intensity with an average of one time purchase, have an average spend of more than $1000, and the average distance of purchase history is more than one year.

I. System Implementation

1. Overview Page

The overview page is a web view that displays information on the results of exploratory data analysis presented in graphical form. Where can users know the contents of the data information held, as seen in **Figure 12**.

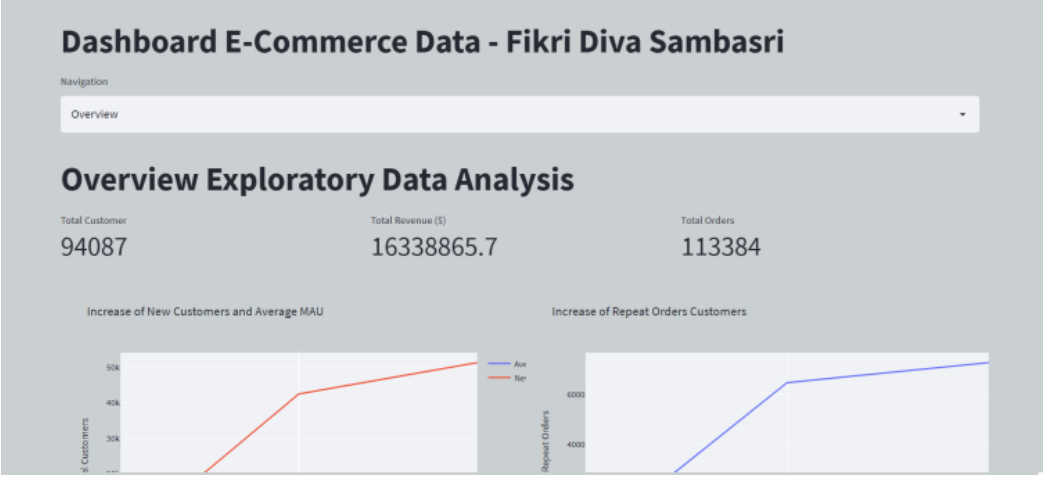

Figure 12. Overview Page View

2. RFM Analysis Page

The RFM analysis page display is a web view that displays the results clustering for each feature, namely recency, frequency, and monetary, as seen in **Figure 13**

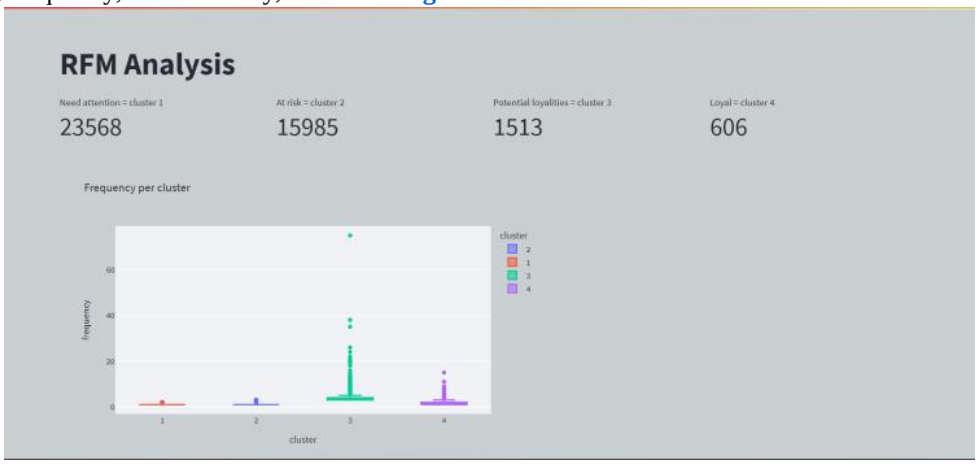

Figure 13. RFM Analysis Page View

3. Report or Cluster Test Page

The report page or cluster test is a page for testing or testing new data; the results of this test will display a graphical visualization as well as the results prediction of clustering from files that have been uploaded on the cluster test page, as seen in **Figure 14**.

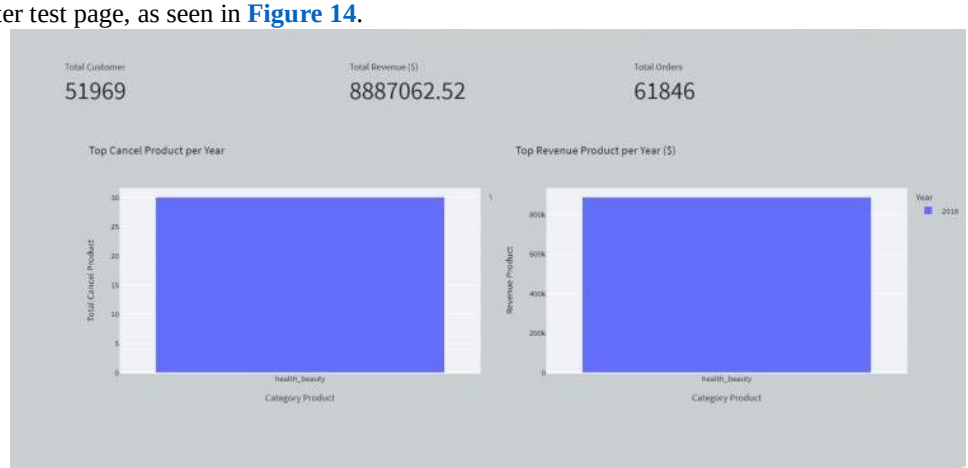

Figure 14. ReportPage View

Conclusion

Based on the research of the Method RFM and K-Means algorithm implementation for customer segmentation in e-commerce with streamlit; the result of feature construction from the E-Commerce Olist dataset feature in the form of the RFM feature has been successfully executed with the findings obtained in the form of values on features frequency at most is the value 1. So, most of the customers only make a one-time purchase of the product throughout the year. The performance of the K-Means clustering algorithm proves that this algorithm can be applied to perform customer segmentation. This is evidenced by the Silhouette score evaluation with a value of 0.470 with a cluster value of 4.

Acknowledgement

We would like to thank Lembaga Penelitian dan Pengabdian Universitas Dian Nuswantoro for Research Grant in Penelitian Dasar Perguruan Tinggi Year 2021-2022 and Center of Excellence Universitas Dian Nuswantoro for providing the guidance and computing resources in this research.

References

[1] S. Palingi and E. C. Limbongan, "Pengaruh internet terhadap industri E-Commerce dan regulasi perlindungan data," in *Seminar Nasional Riset dan Teknologi (SEMNAS RISTEK)*, 2020.

[2] APJII, "Penetrasi & profil perilaku pengguna internet Indonesia," in *Asosiasi Penyelenggara Jasa Internet Indonesia*, 2018.

[3] G. Liu, "An ecommerce recommendation algorithm based on link prediction," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 905–910, Jan. 2022, doi: 10.1016/j.aej.2021.04.081.

[4] S. Monalisa, P. Nadya, and R. Novita, "Analysis for Customer Lifetime Value Categorization with RFM Model," *Procedia Comput Sci*, vol. 161, pp. 834–840, 2019, doi: 10.1016/j.procs.2019.11.190.

[5] Huynh T. K., le H.-D, Nguyen S. V., and Tran H. M., "Applying Peer-to-Peer Networks for Decentralized Customer-to-Customer Ecommerce Model," in *International Conference on Future Data and Security Engineering*, 2020.

[6] Olist, "About Olist," *Available: https://olist.com/pt-br/sobre-nos/*, Jul. 18, 2022.

[7] Latifah N, Widayani A, and Normawati R. A, "Pengaruh perceived usefulness dan trust terhadap kepuasan konsumen pada E-Commerce Shopee," *Jurnal Bisnis dan Manajemen*, vol. 14, p. 84, 2020.

[8] R. Heldt, C. S. Silveira, and F. B. Luce, "Predicting customer value per product: From RFM to RFM/P," *J Bus Res*, vol. 127, pp. 444–453, Apr. 2021, doi: 10.1016/j.jbusres.2019.05.001.

[9] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *Journal of Retailing and Consumer Services*, vol. 61, p. 102566, Jul. 2021, doi: 10.1016/j.jretconser.2021.102566.

[10] V. Holý, O. Sokol, and M. Černý, "Clustering retail products based on customer behaviour," *Appl Soft Comput*, vol. 60, pp. 752–762, Nov. 2017, doi: 10.1016/j.asoc.2017.02.004.

[11] Puspitasari N, Widians J. A, and Setiawan N. B, "Segmentasi pelanggan menggunakan algoritme Bisecting K-Means berdasarkan model Recency, Frequency, dan Monetary (RFM)," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, pp. 78–79, 2020.

[12] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.

[13] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, "Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm," *Appl Soft Comput*, vol. 113, p. 107924, Dec. 2021, doi: 10.1016/j.asoc.2021.107924.

[14] Boentarman M, Rostianingsih S, and Setiawan A, "Penerapan segmentasi pelanggan dengan menggunakan metode K-Means Clustering pada Sistem Customer Relationship Management di PT Titess," *Jurnal Infra*, p. 2, 2021.

[15] Berahmana R, Mohammed F. A, and Chairuang K, "Customer segmentation based on RFM model using K-Means, K-Medoids, and DBSCAN methods," *Lontar Komputer*, vol. 11, pp. 32–38, 2022.