Journal of Science and Applied Engineering (JSAE)
Vol. 8, No. 2, September 2025
DOI : 10.31328/js.v8i2.7302

ISSN Print : 2621-3745
ISSN Online : 2621-3753
(Page.44-55)

CLUSTERING USING K-MEANS ALGORITHM AND RECENCY, FREQUENCY, MONETARY MODEL FOR CUSTOMER SEGMENTATION

Fairuz Jawharah[1,], Eka Dyar Wahyuni[2], Abdul Rezha Efrat Najaf[3]*

[1,2,3]Department of Information System, Universitas Pembangunan Nasional "Veteran" Jawa Timur,Surabaya, Indonesia
**Email: jawharah191201@gmail.com*

Submitted : 1 July 2025; *Revision :* 9 July 2025; *Accepted :* 9 July 2025

ABSTRACT

Understanding customer behavior is a critical component in developing effective and sustainable marketing strategies. This study aims to segment customers of Rayu Manis, a culinary business based in Surabaya, by implementing the Recency, Frequency, Monetary (RFM) model in combination with the K-Means clustering algorithm. Transactional data collected from February 2023 to October 2024 underwent several processing stages, including data preprocessing, RFM scoring, logarithmic transformation, normalization, determination of the optimal number of clusters using the Elbow method, and evaluation using the Silhouette Score and Davies-Bouldin Index. The clustering results revealed that the Sheet Order dataset formed two clusters with a Silhouette Score of 0.51285, while the Sheet Rayu Manis dataset yielded three clusters with a Silhouette Score of 0.656. The resulting segmentation identified groups of loyal, potential, and at-risk customers, providing a data-driven foundation for targeted marketing strategies and supporting strategic decision-making within the context of small and medium-sized enterprises.

Keywords K-Means; RFM Model; customer segmentation; clustering;
Paper type Research paper

INTRODUCTION

After the pandemic, business and social activities gradually returned to normal, creating new opportunities for entrepreneurs and increasing competition among small and medium-sized enterprises (SMEs). In Indonesia, the number of SMEs grew from 3.4 million in 2013 to 4.5 million in 2023 [1].One growing SME in Surabaya is Rayu Manis, which started in 2015 by selling products online and later expanded in 2021 with the launch of Teras Rayu, a café offering a homey atmosphere, gluten-free treats, and engaging activities for all age groups. Despite its unique appeal, Rayu Manis needs to continuously develop effective strategies—such as Customer Relationship Management (CRM)—to retain and expand its customer base. While the business has consistently recorded sales data, it has yet to leverage that data for analytical purposes. So far, the records have served mainly as sales trackers, offering minimal strategic value. This analysis seeks to convert the existing data into actionable insights through customer segmentation, helping Rayu Manis make more informed decisions and strengthen its business strategy.

To support CRM, this study applies customer segmentation using K-Means clustering [2] and the Recency, Frequency, Monetary (RFM) model [3]. K refers to the constant value or the number of clusters to be created [4]. K-Means is a non-hierarchical clustering method that groups similar data into the same cluster, while data with different characteristics are placed in separate clusters [5]. K-Means is widely used due to its simplicity, speed, and adaptability , while the RFM model transforms raw transaction data into meaningful insights that help evaluate customer loyalty [6]. A previous study titled "Recency and Customer Value Segmentation Analysis at AVANA Indonesia Using K-Means Algorithm and RFM Model" explored customer segmentation using the K-Means algorithm and RFM model. The study aimed to identify recency and customer value segments, resulting in nine clusters based on the elbow method. Key segmentations included four recency-based clusters: active, warm, cold, and inactive customers, as well as four recency-frequency clusters: common, ultra-high, low, and high [7].

By grouping customers based on behavioral similarities, the analysis generates segmented profiles, which are visualized through a web-based dashboard. These insights help Rayu Manis

Journal of Science and Applied Engineering (JSAE)
Vol. 8, No. 2, September 2025
DOI : 10.31328/js.v8i2.7302

identify loyal and high-value customers, personalize marketing strategies, and implement targeted loyalty programs—ultimately enhancing customer satisfaction and long-term profitability.

METHOD

This chapter outlines the research workflow based on the CRISP-DM methodology to ensure a structured and effective process. The overall research flow is illustrated in the following Figure 1. Figure 1 represents the CRISP-DM (Cross-Industry Standard Process for Data Mining) model, which is a widely used data mining process model consisting of six phases. CRISP-DM helps identify correlations between various data mining tasks based on user goals, background, and interests, all grounded in data. Each phase is flexible and interconnected, meaning the process is iterative rather than strictly linear [3].

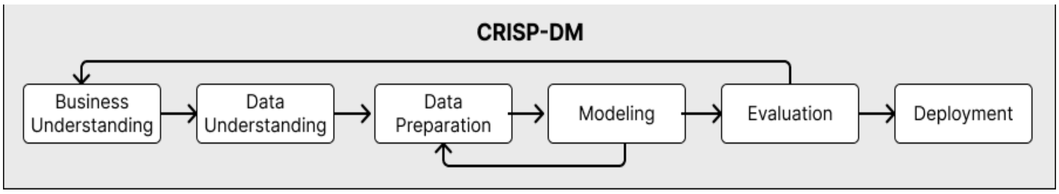


Figure 1. CRISP-DM Method

Business Understanding

This stage requires the researcher to grasp the objectives and requirements of the study from a business perspective. This understanding must then be translated into a clear definition of the data mining problem, which will serve as the foundation for the initial plan to achieve the research objectives.

Data Understanding

After having the clear definition of the data mining problem, it continues to understanding the data for the research. This stage naturally begins by collecting the data then followed by identifying potential quality issues, uncovering initial insights, and detecting interesting subsets that could form the basis for hypotheses about hidden information.

Data Preparation

Since 2023, Rayu Manis has documented its sales using spreadsheets. That year, two files were used: one covering February to October, and another for October to December. In 2024, sales from January to October were recorded in a single file. Each file contains multiple sheets, including Customer, Order, Sales, Product, and Rayu Manis (which records special orders). In total, these datasets consist of 3,686 records. To conduct meaningful analysis from these records, the data must first go through a preparation phase—starting with data selection, followed by data preprocessing, as illustrated in Figure 2.

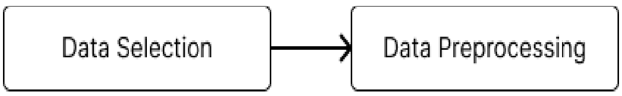


Figure 2. Data Preparation

Not only does this stage involve understanding the data used in the research, but it also requires identifying which data should be selected for analysis. This is followed by data preprocessing, which begins with data cleaning. The cleaning process includes merging datasets, removing unused columns or irrelevant records, encoding data, and handling missing values.

Modeling

There are many models commonly used in data mining, and the choice should be tailored to the specific objectives of the research. Selecting the most appropriate model is crucial to ensure optimal results. One such widely used method is clustering, which has been applied across various disciplines,

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/js.v8i2.7302
(Page.44-55)

research fields, and case studies. As an unsupervised learning technique, clustering groups data without predefined labels—unlike classification, where categories are known in advance. K-Means, in particular, is a non-hierarchical clustering algorithm that organizes data with similar characteristics into the same cluster, while placing dissimilar data into separate clusters[8]. As the name suggests, clustering divides a dataset into several distinct groups or partitions [9].

Evaluation

It is essential to thoroughly evaluate the model and review the procedures used during its development. This ensures that the model aligns with the intended business objectives. One of the key goals is to identify whether any critical business issues have been overlooked.

Deployment

The method ends with deployment. The primary aim of building a model is to enhance insights into the data. However, the knowledge extracted from the model must not only be insightful but also structured and delivered in a format that is practical and accessible for end users. Deployment may range from simple reporting to fully integrating models into decision-making systems, like real-time website adjustments or ongoing marketing analysis.

DISCUSSION

For this research, the process model used is CRISP-DM. The following sections will explain each phase of the CRISP-DM model in detail.

Business Understanding

Since starting the business in 2015, Rayu Manis has practiced basic Customer Relationship Management (CRM) and sales data recording. Initially, they used Paper.id, which provided detailed records on sales and inventory. However, its dashboard was limited—only showing monthly profit and loss—without customer segmentation features necessary for effective CRM. In 2023, alongside two years of operating their café, Teras Rayu, they shifted to recording sales through Google Sheets, documenting customer details, orders, dates, payment, and order types. Despite using these platforms, Rayu Manis has not optimized their CRM, as the data remains raw and unprocessed. Without clear customer segmentation, their outreach through platforms like Instagram and WhatsApp remains too general, leading to less effective CRM strategies.

Data Understanding

As previously described, Rayu Manis has recorded its sales using spreadsheets from 2023 to 2024. In 2023, two files were used (February–October and October–December), while in 2024, a single file covers January–October. Each file contains several sheets, including Customer, Order, Sales, Product, and Rayu Manis (special orders), with the 2024 file also featuring Kue Kering (Ramadan sales) and Pasar Sehat (monthly health market sales).

Data Preparation

a. *Data Selection*

Among all sheets, the most relevant for customer segmentation analysis is the Sheet Order and Sheet Rayu Manis, as it contains comprehensive daily sales data including customer details, products, quantities, and payment methods.

b. *Data Preprocessing*

Microsoft Excel is used as the main tool for this processing.

- Merging Data

As previously explained, the selected data is spread across different files. Therefore, the first step in this stage is to merge all the required data. Figure 3 shows the result of combining the three Order

Sheets from sales files covering February 2023 to October 2024. After merging, the combined Order Sheet consists of 2,477 rows and 14 columns.

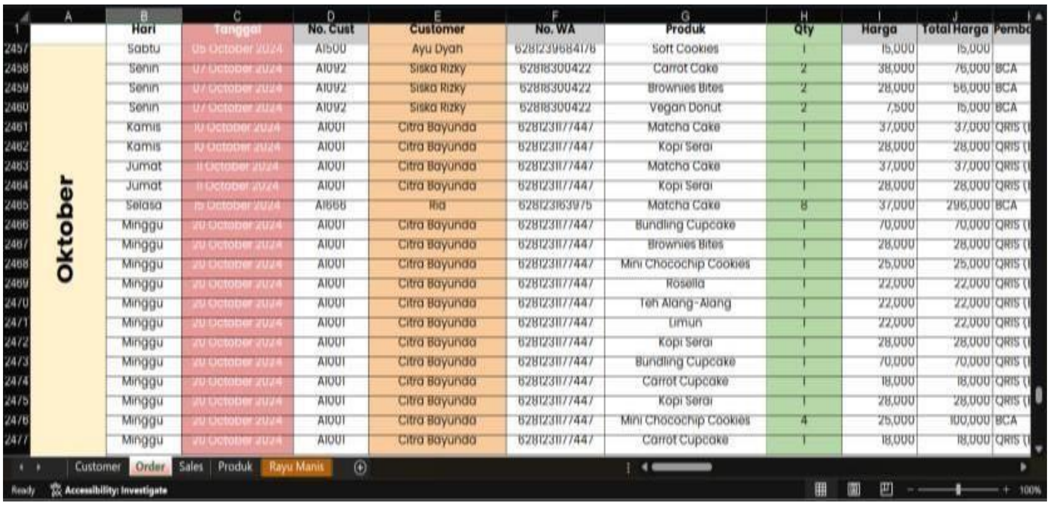


Figure 3. Dataset

- Removing unused column and or irrelevant records

After merging the data, the total number of columns in the Sheet Order is 13. Upon reviewing the analysis requirements and considering the existing dataset, several columns need to be removed because their data will not be used to achieve the research objectives. The column removal process was carried out using Microsoft Excel tools. The details of the columns to be deleted are listed in TABLE 1 below After deleting these columns, the updated number of columns is 11.

TABLE 1. DELETED COLUMN FROM SHEET ORDER

Column	Column Name	Description
A	Bulan	Month when the transaction occurred
M	Ongkir	Description of the total shipping cost of the product
N	Invoice	Information regarding whether an invoice has been created for the customer's order or not

For the Rayu Manis Sheet, there are a total of 14 columns. After deleting some columns, the updated number of columns is 9. The details of the columns to be deleted are listed in TABLE 2 below.

TABLE 2. DELETED COLUMN FROM SHEET RAYU MANIS

Column	Column Name	Description
A	Bulan	Month when the transaction occurred
K	Invoice	Information regarding whether an invoice has been issued for the customer's order
L	Pembayaran	Information on the payment status of the customer's order
M	Pembayaran	Payment method used by the customer
N	Ongkir	Information on the total shipping cost of the product

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/js.v8i2.7302
(Page.44-55)

After removing unnecessary columns, the next step was to delete irrelevant data. Since Rayu Manis is an SME focused on bakery and pastry sales, the study targets analyzing these products. However, because the cafe also sells beverages and other items recorded together, beverage data will be included in the analysis to provide useful customer segmentation insights for sales strategy. Other non-bakery and pastry sales data were removed. There are a total of 44 item types excluded from this research. Therefore, any row of data in the Sheet Order or Sheet Rayu Manis containing any of these 44 item types will also be excluded.

- Encoding data

From February 2023 to October 2024, they initially used all the payment options mentioned earlier. However, starting in 2024 and continuing forward, they have focused on only four payment methods: cash, debit, QRIS, and transfer. In response, the original 11 payment options were consolidated into these 4 using the find and replace feature in Microsoft Excel. This simplification aims to facilitate analysis and make the visualizations easier for Rayu Manis's owner to understand. The details of the changes are listed in TABLE 3 below.

TABLE 3. ENCODED DATA

Original Value	Resulting Change
QRIS (BCA), QRIS (Bank Jatim), QRIS (EDC), QRIS Mandiri	QRIS
Debit (Mandiri), Debit (BCA)	Debit
BCA, Mandiri, Paper ID	Transfer

- Handling missing value

To address missing values in dataset, a cleaning process was carried out using Excel. Specifically, rows containing incomplete or missing entries were identified through the filter function, which enabled efficient detection of null or blank values across key variables. Once identified, these rows were removed using the delete row feature to ensure that the remaining dataset consisted only of complete and reliable records.

Modelling

- *RFM Analysis*

RFM consists of three variables which are Recency, Frequency, and Monetary [10]:

- Recency measures how recently a customer made a purchase. A shorter time since the last transaction indicates a higher recency score.
- Frequency reflects how often a customer makes purchases within a specific period. More frequent transactions result in a higher frequency score.
- Monetary measures the total amount a customer has spent during a given period. Higher spending indicates a higher monetary score.

After the data was cleaned and ready for RFM analysis, the next step was to calculate the first attribute: recency. Recency measures how long it has been since a customer's last purchase. Using the "Tanggal" (Date) column, the number of days between each customer’s most recent transaction and the latest date in the dataset was calculated. This produced a recency value for each of the 414 customers. The next step involved calculating frequency, which reflects how often a customer made transactions. This was done by grouping the data by customer number and counting the number of transaction entries. Lastly, the monetary value—which represents the total spending of each customer—was calculated by summing the "Total Harga" (Total Price) column for each customer. All three metrics—recency, frequency, and monetary—were successfully extracted for a total of 414 customers using Python code. Figure 4 below shows the combined view of the three RFM model attributes for sheet order and Figure 5 for sheet Rayu Manis.

ISSN Print : 2621-3745
ISSN Online : 2621-3753
(Page.44-55)

Vol. 8, No. 2, September 2025
DOI : 10.31328/js.v8i2.7302

No. Cust	Recency	Frequency	Monetary
A1009	17	4.0	1139000
A1010	14	136.0	733000
A1014	1	88.0	56000
A1015	3	60.0	122000
A1021	5	568.0	174000
...	...	...	...

Figure 4. RFM Sheet Order

No. Cust	Recency	Frequency	Monetary
A1001	300	2	700000
A1009	410	1	250000
A1023	216	2	615000
A1046	29	2	500000
A1059	373	2	585000
...	...	...	...

Figure 5. RFM Sheet Rayu Manis

After calculating the values of each RFM attribute (Recency, Frequency, and Monetary), the next step was to examine the data distribution for each of these attributes. This distribution check aimed to assess the skewness in order to determine whether data transformation was needed to achieve more symmetrical distributions. Visualization was done using histograms and distribution curves with the help of matplotlib and seaborn libraries.

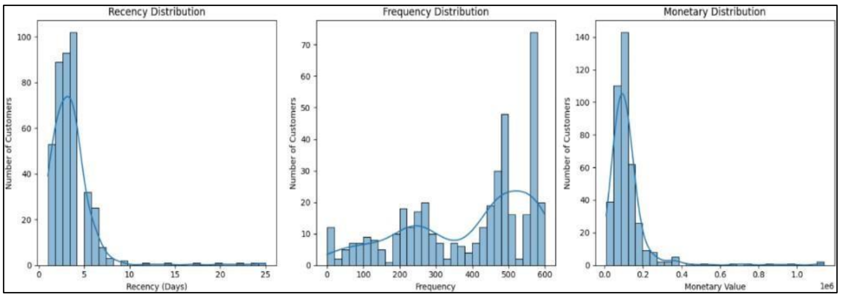


Figure 6. Skewness RFM Sheet Order

The results in Figure 6 showed that Recency and Monetary had high positive skewness, thus requiring log transformation to reduce skewness and normalize their distributions. Meanwhile, Frequency had an acceptable distribution and did not require transformation. After applying log transformation, the distributions of Recency and Monetary became more symmetrical, with significantly reduced skewness values. To ensure all attributes are on the same scale, min-max normalization was applied. This normalization is essential to prevent scale differences from biasing clustering results. The final outcome of this process is a fully transformed and normalized RFM dataset for sheet order as it shown in Figure 7 below.

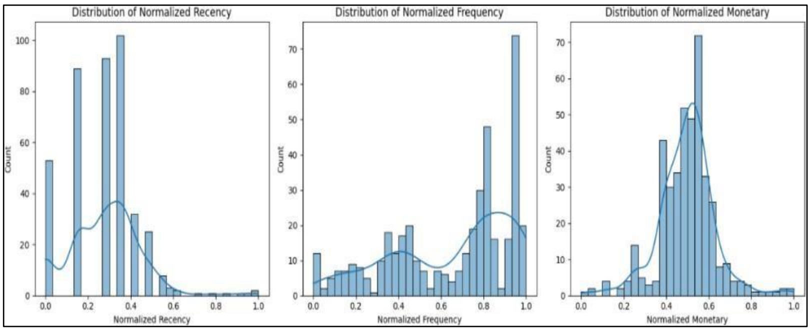


Figure 7. Normalized RFM Sheet Order

Journal of Science and Applied Engineering (JSAE)
Vol. 8, No. 2, September 2025
ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/js.v8i2.7302
(Page.44-55)

Similar to the Sheet Order, after obtaining the values of each RFM attribute, the next step is to examine the data distribution to assess the skewness of the three attributes. Figure 8 illustrates that both frequency and monetary attributes exhibited high positive skewness, indicating the need for log transformation to normalize their distributions. In contrast, the recency attribute displayed a reasonably symmetrical distribution and did not require transformation. Following the log transformation, the distributions of frequency and monetary became more balanced, with notably lower skewness values. To standardize the scale across all attributes, min-max normalization was then applied. The end result of this process is a completely transformed and normalized RFM dataset for the Sheet Rayu Manis, as presented in Figure 9.

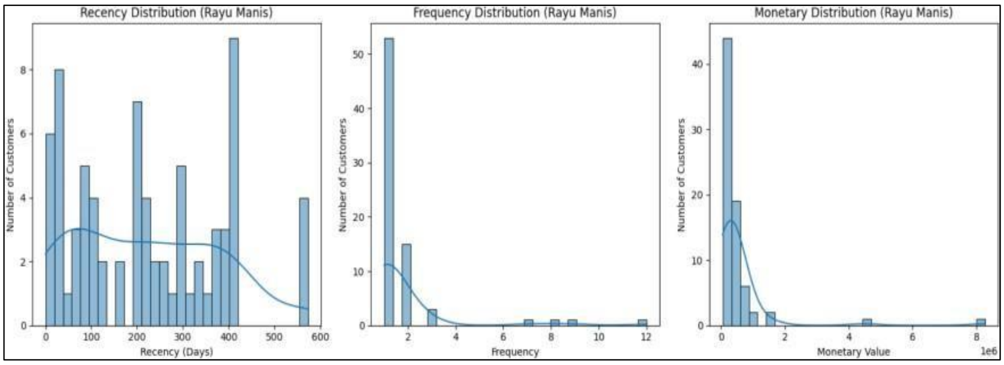


Figure 8. Skewness RFM Sheet Rayu Manis

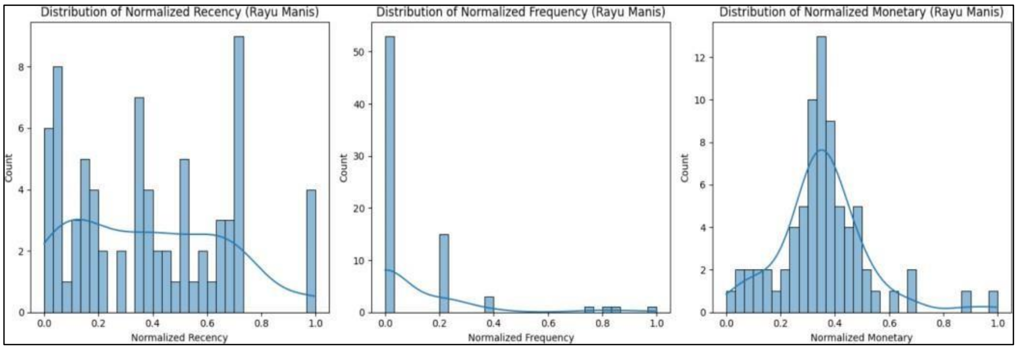
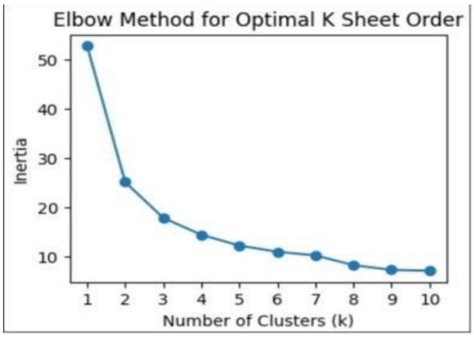


Figure 9. Normalized RFM Sheet Rayu Manis

- Determination of the Optimal K Value

There are several methods for determining the most optimal K value for clustering. In this research, three methods are used: the elbow method, silhouette score, and Davies-Bouldin index. Combining these three methods helps accurately determine the best number of clusters. The first one is elbow method. This process iterates k values from 1 to 10, trains a K-Means model for each k, and stores the SSE values. The results are visualized in the elbow plot to identify the "elbow point," which indicates the optimal cluster count. Subsequently, the decrease in SSE between clusters is calculated to observe significant changes. The result of the elbow method for sheet order is shown in Figure 10 and for sheet rayu manis in Figure 11.

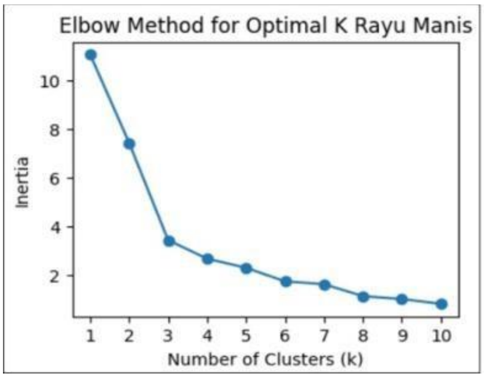


k	SSE	Decrease in SSE
1	52.860311	0.000000
2	25.134755	27.725556
3	17.855986	7.278770
4	14.420669	3.435317
5	12.222273	2.198396
6	10.960727	1.261546
7	10.227924	0.732804
8	8.233074	1.994850
9	7.265898	0.967175
10	7.102250	0.163649

Figure 10. Elbow Result Sheet Order

ISSN Print : 2621-3745
ISSN Online : 2621-3753
(Page.44-55)

Journal of Science and Applied Engineering (JSAE)
Vol. 8, No. 2, September 2025
DOI : 10.31328/js.v8i2.7302



k	SSE	Decrease in SSE
1	11.036558	0.000000
2	7.419217	3.617341
3	3.433717	3.985500
4	2.674241	0.759475
5	2.300904	0.373337
6	1.746502	0.554402
7	1.629956	0.116546
8	1.139067	0.490888
9	1.022381	0.116687
10	0.825139	0.197242

Figure 11. Elbow Result Sheet Rayu Manis

In addition to checking with the elbow method and evaluating the decrease in SSE, silhouette score and davies-bouldin index was also used to find the most optimal K value. The outcome of the elbow method and davies-bouldin index for the sheet order is illustrated in Figure 12 while Figure 13 shows the results for Sheet Rayu Manis.

```
For n_clusters = 2, the average silhouette_score is : 0.5128562822046192
For n_clusters = 3, the average silhouette_score is : 0.49726467345024405
For n_clusters = 4, the average silhouette_score is : 0.37419094835471034
For n_clusters = 5, the average silhouette_score is : 0.34993881887114975
For n_clusters = 6, the average silhouette_score is : 0.3066807536708816
For n_clusters = 7, the average silhouette_score is : 0.31925634764595817
For n_clusters = 8, the average silhouette_score is : 0.3359084959047101
For n_clusters = 9, the average silhouette_score is : 0.3369645051794345
For n_clusters = 10, the average silhouette_score is : 0.3282851381076322
```

```
For n_clusters = 2, the Davies-Bouldin Index is : 0.8316984675233414
For n_clusters = 3, the Davies-Bouldin Index is : 0.8955281279357061
For n_clusters = 4, the Davies-Bouldin Index is : 0.9730197263518698
For n_clusters = 5, the Davies-Bouldin Index is : 1.064752240345408
For n_clusters = 6, the Davies-Bouldin Index is : 1.0706090882266313
For n_clusters = 7, the Davies-Bouldin Index is : 1.1222464253317617
For n_clusters = 8, the Davies-Bouldin Index is : 0.972168623665457
For n_clusters = 9, the Davies-Bouldin Index is : 0.9763850247945322
For n_clusters = 10, the Davies-Bouldin Index is : 1.0457136924546244
```

Figure 12. Silhouette and Davies-Bouldin Sheet Order

```
For n_clusters = 2, the average silhouette_score is : 0.5908814275757891
For n_clusters = 3, the average silhouette_score is : 0.4583641693113535
For n_clusters = 4, the average silhouette_score is : 0.357669682227818
For n_clusters = 5, the average silhouette_score is : 0.3411303035944723
For n_clusters = 6, the average silhouette_score is : 0.42869109028056945
For n_clusters = 7, the average silhouette_score is : 0.4082036151667909
For n_clusters = 8, the average silhouette_score is : 0.44683573392136255
For n_clusters = 9, the average silhouette_score is : 0.44668738576875827
For n_clusters = 10, the average silhouette_score is : 0.4721056135226228
```

```
For n_clusters = 2, the Davies-Bouldin Index is : 0.4770572187696221
For n_clusters = 3, the Davies-Bouldin Index is : 0.693075279405095
For n_clusters = 4, the Davies-Bouldin Index is : 0.9242205610748975
For n_clusters = 5, the Davies-Bouldin Index is : 0.9530103506435317
For n_clusters = 6, the Davies-Bouldin Index is : 0.7351529248656964
For n_clusters = 7, the Davies-Bouldin Index is : 0.7490589691502975
For n_clusters = 8, the Davies-Bouldin Index is : 0.6992342005802203
For n_clusters = 9, the Davies-Bouldin Index is : 0.6628176753327315
For n_clusters = 10, the Davies-Bouldin Index is : 0.6481879471987301
```

Figure 12. Silhouette and Davies Bouldin Sheet Rayu Manis

Based on the evaluation using SSE, Silhouette Score, and Davies-Bouldin Index, the optimal number of clusters for both Sheet Order and Sheet Rayu Manis is k = 2. In both cases, k = 2 shows the steepest SSE drop, the highest silhouette scores (0.512 for Order, 0.591 for Rayu Manis), and the lowest DBI values (0.83 and 0.477 respectively), indicating well-separated and compact clusters. Although minor improvements appear at higher k values, they are not consistent. Thus, k = 2 is the most stable and ideal choice.

Journal of Science and Applied Engineering (JSAE)
Vol. 8, No. 2, September 2025
DOI : 10.31328/js.v8i2.7302
ISSN Print : 2621-3745
ISSN Online : 2621-3753
(Page.44-55)

- Assigning Data to the Nearest Cluster

To calculate the distance between centroids, it is essential to first determine the initial centroids, which are selected randomly. Once the initial centroids are set, each data point is assigned to the nearest cluster based on the Euclidean distance formula.

$$\boldsymbol{D_e} = \sqrt{(\boldsymbol{x_i} - \boldsymbol{s_i})^2 + (\boldsymbol{y_i} - \boldsymbol{t_i})^2}$$

Description:
De = Euclidean Distance
i = number of objects
(x, y) = coordinates of the object
(s, t) = coordinates of the centroid

This is followed by recalculating new centroids and reassigning the data points accordingly. The process is repeated iteratively until the centroids no longer change. For the Sheet Order, the centroid values stabilized after six iterations. The output of the data assigned to clusters based on the nearest centroid is shown in Figure 14 below. For the Sheet Rayu Manis, the centroid values stabilized after four iterations, and the resulting clustered data output is presented in Figure 15.

	No. Cust	Recency_normalized	Frequency_normalized	Monetary_normalized	Cluster
0	A1009	0.856635	0.006678	1.000000	1
1	A1010	0.785553	0.227045	0.911107	1
2	A1014	0.000000	0.146912	0.392430	1
3	A1015	0.270238	0.100167	0.549471	1
4	A1021	0.428317	0.948247	0.621074	0

Figure 13. Assigned Data to Cluster Sheet Order

	No. Cust	Recency_normalized	Frequency_normalized	Monetary_normalized	Cluster
0	A1001	0.523560	0.216618	0.516735	0
1	A1009	0.715532	0.000000	0.315132	0
2	A1023	0.376963	0.216618	0.491387	1
3	A1046	0.050611	0.216618	0.450853	1
4	A1059	0.650960	0.216618	0.481595	0

Figure 14. Assigned Data to Cluster Sheet Rayu Manis

While both datasets resulted in two distinct clusters, the patterns and characteristics of the customer segments differed across them. The following sections elaborate on the clustering results for each dataset.

- Sheet Order Dataset

The first dataset, Sheet Order, contains 414 customer entries. Clustering with the K-Means algorithm yielded two optimal clusters, supported by a silhouette score of 0.51285, which indicates a good level of cluster separation. In this analysis, Cluster 0 was identified as representing Loyal Customers. Customers in this group showed consistently high purchase frequency, large transaction values, and relatively recent transaction dates. These characteristics reflect strong customer engagement and loyalty, suggesting that this group may respond well to retention and reward strategies.

In contrast, Cluster 1 consisted of customers with low purchase frequency and a long duration since their last transaction. This group was categorized as At-Risk Customers, indicating a decline or gap in their purchasing activity. The clear behavioral distinction between these two clusters allows for more targeted customer relationship management, such as reactivation campaigns for the at-risk group and loyalty programs for the high-value segment.

ISSN Print : 2621-3745
ISSN Online : 2621-3753
(Page.44-55)

Journal of Science and Applied Engineering (JSAE)
Vol. 8, No. 2, September 2025
DOI : 10.31328/js.v8i2.7302

- Sheet Rayu Manis Dataset

The second dataset, Sheet Rayu Manis, contains 75 customer entries. The K-Means clustering process similarly resulted in two clusters, with a silhouette score of 0.40095, which, while slightly lower than the previous dataset, still reflects an acceptable clustering outcome. However, the pattern observed in this dataset differed from that of Sheet Order. Here, Cluster 0 represented the At-Risk Customers, who demonstrated lower purchasing frequency and a longer time since their last purchase. Meanwhile, Cluster 1 contained customers categorized as Loyal Customers, distinguished by more frequent and recent purchases.

This reversal in cluster composition suggests that customer behavior in the Rayu Manis dataset diverges from that in Sheet Order. Several factors may contribute to this difference, including variations in data collection methods, differences in the time periods represented by each dataset, or distinct customer segmentation approaches applied within each business context. These findings underscore the importance of context-specific analysis when interpreting clustering results and highlight the need to align marketing strategies with the behavioral characteristics of each customer segment.

Based on the characteristics of each customer segment, different marketing strategies can be formulated to optimize engagement and retention. For the Loyal Customers segment, recommended strategies include the implementation of loyalty programs, exclusive offers, personalized communication, and the promotion of high-value products. These approaches aim to strengthen the relationship with loyal customers and encourage continued purchasing behavior. In contrast, for the At-Risk Customers segment, more proactive efforts are required to regain their interest and re-establish engagement. Suitable strategies may involve reactivation programs such as special discounts, targeted and engaging marketing campaigns, the introduction of new product offerings, and intensified communication through social media platforms. These actions are intended to re-capture the attention of disengaged customers and encourage them to return to active purchasing behavior.

Evaluation

After the data from each sheet has been assigned to clusters based on the nearest centroid, the study continues by evaluating the clustering results using the silhouette coefficient method. The clustering of the Sheet Order produced two clusters with a silhouette score of 0.51285, while the clustering of Sheet Rayu Manis also resulted in two clusters with a silhouette score of 0.40095. These scores indicate that the clusters are fairly well-formed, though some overlap between clusters still exists, suggesting that further optimization may be necessary.

Deployment

The final stage of this research is deployment, where the clustered data is visualized through a web-based interface. Using the Flask framework, the clustering results are displayed on a website accessible via https://iruz19.pythonanywhere.com/. The visualizations include various charts such as 3D and 2D scatter plots as well as pie charts, utilizing Chart.js and Plotly libraries. Figure 16 below shows a portion of the main dashboard display, which includes all charts representing the clustered sales data of Rayu Manis from both the Sheet Order and Sheet Rayu Manis. This page also features a pie chart illustrating the percentage of data in each cluster, along with a table showing the number of data points in each cluster.

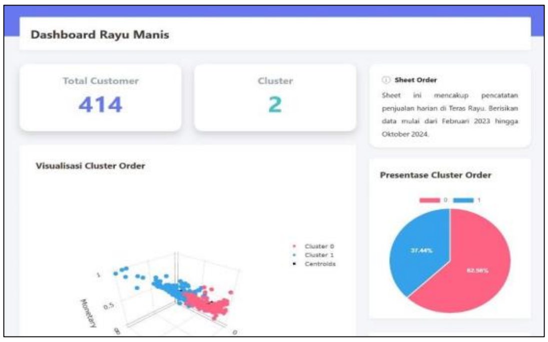


Figure 15. Dashboard Web Visualization

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/js.v8i2.7302
(Page.44-55)

Here are the detailed results of the clustering that has been carried out.

a. Sheet Order

The visualization results for the Sheet Order clustering show that out of 414 data points, 62.6% fall into Cluster 0 (Red) and 37.4% into Cluster 1 (Blue). Cluster 0 is characterized by customers with high frequency and monetary values and low recency, indicating frequent and recent transactions with high spending. In contrast, Cluster 1 includes customers with lower frequency, high recency, and generally lower monetary values. These characteristics are supported by the 3D and 2D scatter plots, where Cluster 0 consistently represents more active and valuable customers, while Cluster 1 shows a more varied distribution with generally lower engagement and value.

b. Sheet Rayu Manis

The clustering results for the Sheet Rayu Manis data show a total of 75 entries, with Cluster 0 (Red) containing 45.3% (34 data points) and Cluster 1 (Blue) comprising 54.7% (41 data points). Based on the 3D scatter plot, Cluster 0 consists of customers with moderate frequency and monetary values, but higher recency, indicating they haven't transacted recently. In contrast, Cluster 1 includes customers with low recency and higher frequency and monetary values, suggesting they are more active and valuable. The 2D scatter plots further support this: Cluster 0 customers tend to have higher recency and lower frequency and spending, while Cluster 1 customers are characterized by low recency and high frequency and spending. This indicates that Cluster 1 represents the more engaged and profitable customer segment.

CONCLUSION

The clustering analysis on Rayu Manis sales data using K-Means and the RFM model resulted in two clusters for each dataset. In Sheet Order, Cluster 0 consists of Loyal Customers with high frequency, high monetary value, and low recency—indicating active and valuable buyers. Cluster 1 includes At-Risk Customers with lower transaction frequency and higher recency. Similarly, Sheet Rayu Manis produced Cluster 0 as At-Risk Customers with low activity and high recency, while Cluster 1 identified Loyal Customers who recently made high-value, frequent purchases. Silhouette scores (0.51 and 0.40) show moderate clustering quality with potential for refinement.

To support the interpretation of clustering results and assist the owner of Rayu Manis in understanding customer segmentation, a series of visualizations were employed. The visual representations provided clear insights into the position and characteristics of each cluster, thereby informing more targeted strategic decisions. For the Loyal Customers segment, recommended strategies include loyalty programs, exclusive offers, personalized communication, and the promotion of high-value products. For the At-Risk Customers, suitable approaches involve reactivation efforts such as special discounts, engaging marketing campaigns, new product offerings, and intensified communication via social media. Through a combination of data-driven analysis and visual interpretation, the segmentation outcomes are expected to serve as a foundation for designing more effective and sustainable customer retention and marketing strategies for Rayu Manis.

REFERENCES

[1] Badan Pusat Statistik, "Jumlah Perusahaan Industri Skala Mikro dan Kecil Menurut Provinsi (Unit), 2013," Tabel Statistik. Accessed: Jun. 24, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NDQwIzI=/jumlah-perusahaan-industri-skala-mikro-dan-kecil-menurut-provinsi.html

[2] P. Putri, "Analisa Nilai K terhadap Pengelompokkan Penjualan Tiket pada K-Means dengan Metode Elbow dan Silhouette Elbow and Silhouette Methods for K Value Analysis of Ticket Sales Grouping," vol. 13, pp. 28–38, 2024.

[3] A. Febri, N. Ningsih, and J. Lemantara, "Aplikasi Analisis Segmentasi Pelanggan untuk Menentukan Strategi Pemasaran Menggunakan Kombinasi Metode K-Means dan Model RFM," Sistemasi, vol. 10, no. 1, p. 152, 2021, doi: 10.32520/stmsi.v10i1.1123.

[4] I. Arfyanti, T. Bustomi, and I. Haristyawan, "Implementasi Data Mining dengan Menerapkan Algoritma K-Means Clustering untuk Memberikan Rekomendasi Jurusan Kuliah Bagi Mahasiswa Baru," vol. 8, no. April, pp. 988–1000, 2024, doi: 10.30865/mib.v8i2.7429.

[5] S. W. Hadi, M. F. Julianto, S. Rahmatullah, and W. Gata, "Analisa Cluster Aplikasi Pada App Store Dengan Menggunakan Metode K-Means," Bianglala Informatika, vol. 8, no. 2, pp. 86–90, 2020, doi: 10.31294/bi.v8i2.8191.

[6] G. P. Hafidz and R. U. Muslimah, "Pengaruh Kualitas Layanan, Citra Merek, Kepercayaan Pelanggan Dan Kepuasan Pelanggan Terhadap Loyalitas Pelanggan Produk Herbalife," Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA), vol. 7, no. 1, pp. 253–274, 2023, doi: 10.31955/mea.v7i1.2912.

[7] M. Jordy, A. Triayudi, and I. D. Sholihati, "Analisis Segmentasi Recency dan Customer Value Pada AVANA Indonesia Dengan Algoritma K-Means dan Model RFM (Recency, Frequency and Monetary)," Journal of Information System Research (JOSH), vol. 4, no. 2, pp. 579–589, 2023, doi: 10.47065/josh.v4i2.2950.

[8] M. T. I. Rahmayani, "Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means," Jurnal Inovasi Teknik Informatika, vol. 1, no. 2, pp. 40–44, 2018.

[9] L. Pamungkas, N. A. Dewi, and N. A. Putri, "Classification of Student Grade Data Using the K-Means Clustering Method," Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 13, no. 1, pp. 86–91, 2024, doi: 10.32736/sisfokom.v13i1.1983.

[10] S. Sharyanto and D. Lestari, "Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce," JURIKOM (Jurnal Riset Komputer), vol. 9, no. 4, p. 866, 2022, doi: 10.30865/jurikom.v9i4.4525.