International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I8P110

Volume 74 Issue 8, 154-164, August 2026

Original Article

Customer Segmentation using Data Mining Technique Method Case: Sarinah Department Store

Frisca Adiesthy Irdiani[1], Tuga Mauritsius[2]

[1,2]*Information System Management Department, Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia.*

[1]*Corresponding Author : frisca.irdiani@binus.ac.id*

Received: 22 September 2025 Revised: 05 July 2026 Accepted: 11 July 2026 Published: 29 August 2026

Abstract - *In this digitized era, implementing customer transaction data is essential, particularly in the case of department stores such as Sarinah, which have a wide range of customers whose purchasing habits vary widely. Sarinah records large transaction data, but most is used only to present sales reports and is rarely explored in-depth to understand customer behavior, customer value, or customer loyalty. As a result, marketing at Sarinah continues to be rather wide and undiversified and has a limited impact on day-to-day business decisions. This study aims to apply advanced data analysis techniques to segment the customers of Sarinah based on their transactions from 1st March 2022 to 9th September 2025. Data cleaning and normalization are performed before this procedure. Then, feature extraction is done with Recency, Frequency, and Monetary (RFM). After that, the K-Means, K-Medoids, Fuzzy C-Means, and DBSCAN clustering algorithms are used. Then the resulting clusters are evaluated by using the Davies-Bouldin Index (DBI). The results suggest that the best cluster is formed by the 4 clusters in K-Means with a DBI score of 0,571703631365945. There are four distinct types of customers: new, occasional, inactive/at-risk, and loyal. These segments have differing characteristics with respect to their shopping behavior, values, and possibilities for loyalty. This study provides a customer segmentation methodology that can be applied generally to the retailing sector to obtain focused marketing, customer retention, and better management strategies for the Sarinah Department Store.*

Keywords - *Customer segmentation, Clustering algorithms, Data mining, Sarinah club, Sarinah department store.*

1. Introduction

The retail industry is undergoing a rapid change as customers' behaviour, technologies, and competition are changing. Department stores are no longer just competing on products and price; they're also competing based on their knowledge of the customers and how they utilize transaction data. The ability to analyse customer transactions is vital for more informed business decisions that are sustainable and effective.

Sarinah Department Store has a large volume of customer transaction data that continues to grow in line with daily sales activity. But this data is not currently analysed to learn more about customers' behaviour, just used to know the total sales and product performance. Consequently, management does not have a complete view of the heterogeneity of customers, differences in the value of customers, and the patterns of customer loyalty over time.

The direct impact of this limited use of transaction data is on Sarinah Department Store's marketing strategy. The programs and offers are generally mass-marketing without taking into account other customer properties, such as high-volume customers, potential customers, and customers who are at risk of churning. This resulted in poor marketing costs, poor promotional effectiveness and lost opportunity to add value to customers in the long term.

To make a business successful, retailers need to understand their customers' behaviours in the retail digital environment. Sarinah's goal is to create a better loyalty program, using behaviour segmentation. This helps Sarinah learn more about the needs and preferences of each customer segment to develop rewards and promotions. The loyalty programmes will be more meaningful and appealing, making customers want to use loyalty points and stay loyal.

According to 2024 data from Sarinah Department Store, the number of loyalty points issued to customers is significantly higher than the points that are actually redeemed. Sarinah distributed over 1,3 billion loyalty points during Sarinah's Digital Transformation from March 2022 to September 2025; however, the average rate of point redemption by customers was around 3,58%, as presented in Table 1.

Table 1. Data loyalty points distribution vs redemption

Month	Year	Distributed	Redeemed
March	2022	18,486,337	-
April	2022	50,541,043	-
May	2022	51,918,490	-
June	2022	40,119,138	-
July	2022	26,377,709	-
August	2022	24,213,411	-
September	2022	17,585,564	-
October	2022	18,606,874	-
November	2022	42,284,235	-
December	2022	28,905,078	-
January	2023	34,893,496	-
February	2023	47,544,973	-
March	2023	20,651,897	-
April	2023	191,686,977	-
May	2023	197,119,575	-
June	2023	197,160,411	-
July	2023	22,935,041	-
August	2023	22,964,815	3,116,120
September	2023	18,874,043	6,406,738
October	2023	18,759,904	2,269,292
November	2023	27,564,086	-
December	2023	18,985,641	1,321,003
January	2024	17,719,379	-
February	2024	19,479,399	42,000
March	2024	24,176,503	-
April	2024	19,625,565	-
May	2024	16,516,469	-
June	2024	9,732,136	279,000
July	2024	13,540,595	-
August	2024	18,411,763	36,431
September	2024	13,073,501	100,100
October	2024	11,691,727	-
November	2024	12,399,665	18,904
December	2024	15,601,986	18,000
January	2025	47,998,343	-
February	2025	17,472,972	-
March	2025	41,435,037	339,104
April	2025	26,925,114	94,799
May	2025	41,011,431	-
June	2025	58,183,855	43,000
July	2025	60,018,938	200,000
Agust	2025	47,781,029	140,000
September	2025	32,781,906	273,500
October	2025	38,172,785	215,700
November	2025	40,967,820	8,991,463
December	2025	63,209,839	56,167,687

This low redemption rate suggests that most customers have yet to be rewarded with the value they should be receiving out of the loyalty programs offered. This situation had led to an incongruence between the reward systems, promotions, and customers' actual behaviour. If there is a lack of understanding about customers' characteristics and preferences, loyalty programmes are a marketing cost that have little effect on customer loyalty or frequency of purchase.

If this trend persists, then Sarinah Department Store will have a higher chance of losing their customers because there is a low chance of engagement with the loyalty program. The continued impact of these problems is not only seen in the performance of marketing campaigns, which tend to be less optimal and have high costs, but also a slowdown in overall sales performance.

This article points out that the Sarinah Department Store problem is not only about customer transaction data, it is about the lack of systematic means to process and use this data to understand customer shopping behavior. Hence, it is essential to have a customer segmentation on a daily basis and correlate the segmentation with more actionable marketing and loyalty programs. This will hopefully yield better business decision-making that is customer-centric.

1.1. Research Gap

Many studies have been conducted on customer segmentation using data mining techniques and their contribution to data-driven marketing strategies. Recency, Frequency, and Monetary (RFM) analysis has been used in various studies with different clustering algorithms, with success in customer segmentation. The segmentation results, however, were highly dependent on the characteristics of the data set and the clustering technique employed.

Past studies indicate that the majority of studies have been performed in an e-commerce setting or with public data. The purchase behaviour of department stores is therefore relatively under-researched due to its complex products, re-purchase behaviour, and different loyalty initiatives. Moreover, many studies involve a comparison of one or two clustering algorithms and only use a few basic evaluation metrics.

Therefore, if customer segmentation is established, the quality of the segmentation results is not always adequately validated and is not directly connected to a tangible business. Such constraints are even more relevant in Sarinah Department Store, where there is a lack of customer segmentation that matches customer shopping behaviour, resulting in various customer segments that are not properly recognized, consequently leading to low customer loyalty, thus causing a decline in customer participation in marketing campaigns. If there is no understanding of the difference in characteristics and values of these customer segments, then promotions will most likely not work.

From this research gap, this study hopes to use RFM analysis and four clustering methods: K-Means, K-Medoids, Fuzzy C-Means, and DBSCAN to segment the customers at Sarina Department Store. These algorithms have been chosen because they have different behaviours when dealing with data distribution and clustering behaviour, which are typical characteristics of retail transaction data.

The novelty of this research lies not only in the application of different clustering techniques but also in the attempt to measure the quality of the respective segmentation results and to understand the quality measures in the context of business decision -making.

Hence, this study is expected to yield a customer segmentation technically and practically relevant to the design of promotional campaigns and the management of Sarinah Department Store's loyalty, that meet the needs, preferences, and behaviours of each Sarinah Department Store customer segment.

2. Literature Review

According to (Khumaidi et al., 2023), customer segmentation refers to a vital technique that views customer purchasing behaviour based on transactional data. The purpose of this study was to come up with the RFM-AR model along with the K-Means algorithm for clustering the customers based on the recency, frequency, and monetary values.

The findings indicated that the RFM method successfully identified high-value customer clusters. However, the assessments of the clusters' quality were conducted internally and did not take the business implications into account.

(Christy et al., 2021) and (Anitha & Patil, 2022) pointed out that RFM is the most adopted approach for segmentation since it is simple and can capture the purchasing history of customers. Both studies applied the K-Means algorithm to segment the customers, but K-Means still has its limitations, such as dealing with outliers and complex data distribution variations, especially in the retail sector, where the transaction characteristics are heterogeneous.

To address the weakness of K-Means, Fahrudin and (Fahrudin & Rindiyani, 2024) compared K-Means and K-Medoids in RFM-based customer segmentation. This study showed that K-Medoids is more robust against noise and outliers than K-Means. However, this study only compared two clustering algorithms and did not explore other methods for non-linear data structures.

A multi-algorithm approach was explored by Idowu et al., 2019), who combined K-Means, Hierarchical Clustering, and Fuzzy C-Means in customer segmentation. In Fuzzy C-Means, a single customer allows to have membership levels in more than one cluster. However, this study did not comprehensively evaluate cluster quality and did not link segmentation results to concrete marketing strategies.

Other studies by (Deng & Gao, 2020) and (Zhang et al., 2021) applied K-Means and optimization algorithms to an e-commerce context. The results showed improved customer segmentation accuracy, but the e-commerce context has different transaction characteristics than department stores, particularly in loyalty patterns, frequency of physical visits, and use of point-based loyalty programs.

In the telecommunications industry, (Rachmahwati et al., 2022) used K-Means to cluster customers based on transaction value. Although successful in forming customer segments, this study did not evaluate cluster stability and did not provide specific business recommendations for each segment.

Alternative approaches, such as Fuzzy C-Means and DBSCAN, are used in non-e-commerce domains. (Setiawan et al., 2023) and (Andriyani & Puspitarani, 2022) used DBSCAN to handle data with noise and irregular cluster shapes. Although this method excels in handling data complexity, its application to department store transaction data remains relatively limited.

Moreover, (Hamidi & Haghi, 2024) and (Fan, 2023) highlight the significance of using data mining and machine learning to analyze customer behaviour and boost the customer lifetime value. However, most of these studies still focus on prediction or classification, rather than on comprehensively validating the quality of customer segmentation.

(Ros et al., 2023) point out that the assessment of clustering results. One of the most popular internal index used to measure intra-cluster homogeneity and inter-cluster separation is the Davies-Bouldin Index (DBI). There have been many studies in the past that have just employed one evaluation criterion and not linked it with business interpretation.

From the literature review, it can be concluded that the use of RFM and clustering techniques in the field of customer segmentation has gained wide acceptance yet most of the studies are limited to e-commerce or public datasets, a small number of clustering algorithms, and have not yet related the segmentation results to real business issues such as the effectiveness of loyalty programs and the use of customer information for decision making in marketing.

This study was undertaken based on these findings, which led to the idea of customer segmentation of Sarinah Department Store. This study used the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology as the main, structured and systematic process to analyze the data, which is in line with the business objective. In this way, the research can not only be done in the modelling stage, but also in the interpretation stage, where it is necessary to take a look at the results of the research in the management of marketing strategies and customer loyalty programs. Therefore, it is hoped that the research results will be useful both in the academic field and for practical use in implementing data-driven decision-making in the retail sector of the department store.

3. Proposed Methodology

To conduct customer segmentation based on transaction data to support more effective marketing decision-making and loyalty program management at Sarinah Department Store, this research uses a data mining approach combining Recency, Frequency, and Monetary (RFM) analysis and several clustering algorithms.

The quality of the segmentation results is evaluated quantitatively using the Davies-Bouldin Index (DBI) metric.

Due to its domain-independent, systematic nature, its ability to bridge business objectives with technical data analysis processes, and ensure all research stages are clearly documented, this research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) as the methodological framework.

The stages of CRISP-DM applied in this study are Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment, as illustrated in Figure 1.

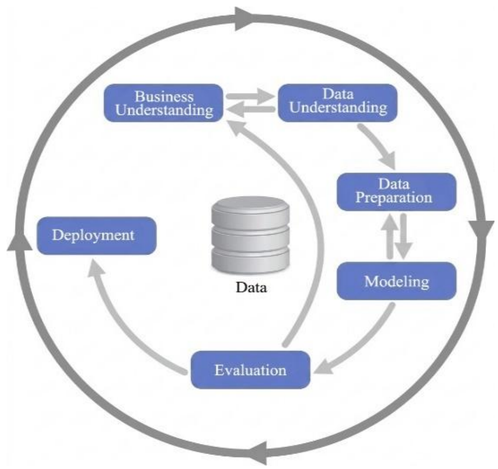

Fig. 1 CRISP-DM stages

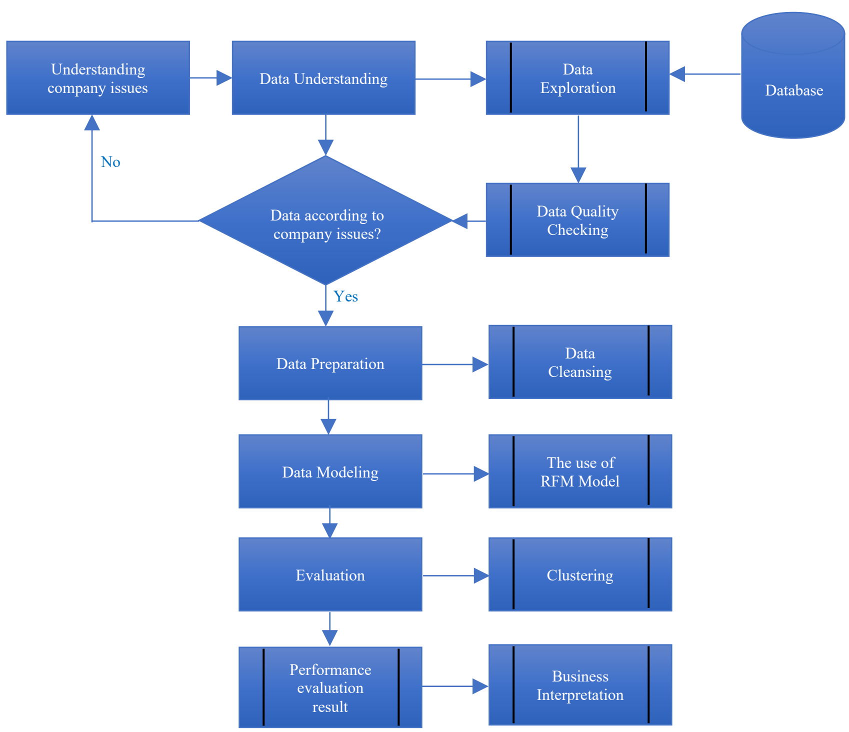

Fig. 2 Research workflow

In addition, the figure shown in this research is an adaptation of CROSP-DM used specifically in the context of Sarinah Department Store. The diagram illustrates the technical implementation of each stage of CRISP-DM, from comprehension of the company's business problems, through exploration and quality checks of transactional data, data preparation processes, RFM-based modelling and clustering algorithms, to the evaluation of the model performance and presentation of results in order to facilitate the business decision-making.

3.1. Business Understanding

The purpose of the Business Understanding stage is to pinpoint the issues that Sarinah Department Store has in its business and to clearly state the goals of the analysis. In Indonesia, Sarinah is one of the retailers that has introduced a point-based loyalty program to keep and enhance customer value. Based on the internal data, though, the loyalty point redemption rate is comparatively low compared to the number of points distributed, suggesting a lack of complete alignment with customers' preferences and shopping behavior with the promotional strategy implemented. This is a case where it is necessary to have a better understanding of the heterogeneity of customers.

Thus, the first aim of the research is to develop more accurate and business-relevant customer segments, using customer transaction history. This helps in creating more effective and targeted marketing campaigns, boosting customer satisfaction and engagement, and enabling data-driven marketing decisions.

3.2. Data Understanding

During the Data Understanding phase, the researcher took a close look at the dataset's details and pinpointed its key features. This dataset came directly from actual transaction records of Sarinah Department Store's Loyalty Management System. This contained the following:

- Data Source: Sarinah Loyalty Management System
- Data Period: March 1, 2022 – September 9, 2025
- Initial Number of Transactions: 63,340 transactions
- Number of Transactions After Cleaning: 32,839 transactions

The primary data items in the study are quite simple, there are:

- Loyalty ID
- Transaction Date
- Transaction Value (Monetary)
- Number of Transactions
- Points Earned
- Points Redeemed

With these attributes, the researcher can compute Recency, Frequency, and Monetary (RFM) values. This is the basis for creating segments of customers.

3.3. Data Preparation

The primary purpose in the Data Preparation phase is to identify missing data and ensure the dataset is uncorrupted, thus ensuring it can accurately be used for clustering analysis. The process of data preparation takes place in multiple stages as depicted in Figure 3.

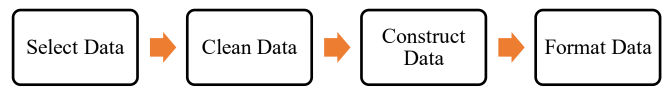

Fig. 3 Data preparation stages

3.3.1. Select Data

The researchers will be concentrating on data that is directly relevant to these objectives, such as loyalty membership transaction data between March 1, 2022 and September 9, 2025.

The researchers also left out non-member transactions since they did not have the necessary details to calculate Recency, Frequency, and Monetary (RFM) values. In this manner, only the information related to the problem of the business and the segmentation goals advanced in the analysis.

3.3.2. Clean Data

Data cleaning was performed to enhance data quality by resolving inconsistencies and errors. The data cleaning process consisted of a number of technical procedures, such as:

- Dealing with missing values. If there are records missing intersection values or transaction times, these are excluded as they would not allow the accurate calculation of RFM.
- Eliminating duplicate data that shares the same Customer ID, transaction date, and transaction values.

3.3.3. Construct Data

The derived data used for the segmentation analysis was developed in the Construct Data stage. From the cleaned transaction data, the following values were calculated:

- Recency, which is the time interval between the customer's last transaction and the research reference date
- Frequency, which is the number of transactions made by the customer during the observation period
- Monetary, which is the total value of the customer's transactions

These three attributes from the RFM model are used as the main features in the clustering process.

3.3.4. Format Data

The final stage is Data Formatting, which aims to standardize the structure and format of the dataset before use in the modelling phase. Data previously collected comes in Comma-Separated Values (CSV) format, and standardizing the format is important before loading the data into a data processing software, which is RStudio.

3.4. Modeling

The Modeling stage is the core phase of the analysis, where the normalized RFM (Recency, Frequency, and Monetary) dataset is used as the basis for developing a customer segmentation model. At this stage, the modeling process focuses not only on technical cluster formation but also on establishing an analytically meaningful and business-relevant customer segment structure.

The process starts by applying the RFM model, which represents each customer as a three-dimensional numeric vector (R, F, M)

- Recency: how much time has passed since the customer's last transaction
- Frequency: how often the customer makes transactions
- Monetary: the total amount of the customer's spending

This vector provides a representation of a feature space to form the groundwork for clustering. The next step is clustering itself; with 4 different algorithms, each of them has its strengths:

- K-Means: Chosen as the baseline model, as it is very useful for large datasets and particularly suited for forming clusters based on distance.
- K-Medoids: Used to overcome the weakness of K-Means in handling outliers, because the cluster centers (medoids) are determined by the actual objects in the data. This makes the model more stable against extreme values in customer transaction data.
- Fuzzy C-Means: Chosen because it allows each customer to have a degree of membership in more than one cluster, thus representing the reality of retail customer behavior, which is not always categorical or exclusive between segments.
- DBSCAN: Finds clusters of any shape and automatically identifies noise. This is handy for retail data, which often isn't evenly spread out and has areas of varying density.

For clustering algorithms that require deciding on the number of clusters, the process relies on:

- The Elbow Method, to spot where there's a clear drop-off in intra-cluster variance.
- The Davies-Bouldin Index (DBI), which measures how well-separated and tight the clusters are.

The configuration showing the least DBI is selected by the researchers as the optimum cluster structure, and it is assumed to be the best segmentation. To sum up, the modelling just does not produce chunks of math, it will create a clear, stable, measurable basis for customer segmentation.

3.5. Evaluation

During the Evaluation phase, both numbers and meaning are addressed. The researcher will first examine if the data is statistically valid and whether the data is of business interest. The primary tool for the quantitative side is referred to as the Davies-Bouldin Index (DBI).

- The lower the DBI, the better the clustering.
- This illustrates that there is high similarity and distinctiveness among members of a cluster.
- Using DBI, researchers can carry out an objective comparison of the quality of different algorithms and select the optimal algorithm depending on the actual data.

Researchers then further interpret the data qualitatively through a focus on:

- Below-average RFM values for each cluster.
- To check the size of every cluster with respect to the Customer.
- Discuss purchasing habits, including purchase frequency, brand loyalty, and revenue analysis.
- The aim is to establish simple, easily understood profiles of each of the customer groups.

It is not sufficient to obtain great metrics. However, these findings are required before the researchers make any sense of them for the business. The process can not end till quantitative scores are taken. It is the process of making the customer segments actionable for managers, which is a major reason.

3.6. Business Interpretation

Business Interpretation get it all tied off and returns to the link strategy analysis. The researchers' explanations of each segment in the activities and aspirations of Sarinah Department Store. This includes:

- List some of the strategic customer types that can be identified:
    - High-value customers
    - Loyal shoppers
    - People who are likely to not repeat orders
    - People who will purchase once in a while or one time
- Creating ideas for promoting each segment, for example:
    - Personalizing campaigns
    - Allocate rewards and points for tactics
    - To differentiate loyalty programs
    - Making things better to ensure customer retention

- Creating valuable insights to support managers in creating data-driven marketing strategies.

The research does not just classify customers. It also gives Sarinah Department Store the insights needed to guide their digital retail transformation and use segmentation as a strategic advantage.

4. Results and Discussion

4.1. Business Understanding

To raise sales, Sarinah needs to understand its customers' shopping behavior, which means understanding the data from the Sarinah Loyalty Management System more deeply.

By choosing the most effective clustering methods and defining customer profiling based on customers' behaviors, Sarinah can create data-driven managerial implications to improve how they engage with their shoppers to remain loyal.

4.2. Data Understanding

The dataset used for this research is Sarinah Club members' transaction records from the Sarinah Loyalty Management System. The dataset will be analyzed using 3 (three) categories:

- Recency: Time since last transaction
- Frequency: Total number of transactions
- Monetary: Total transaction amount

4.3. Data Preparation

The analysis will focus on transaction data collected from Sarinah Club members' transaction records from the Sarinah Loyalty Management System for over 18 months, from March 1, 2022, to September 9, 2025.

The dataset consists of 63.340 raw transaction records in Comma-Separated Values (CSV) format. In the first data assessment, several quality issues were identified, including format inconsistencies, data redundancy, and missing values. Through data cleansing, the dataset was reduced from 63.340 to 32.839 data records.

Table 2. Data cleansing transaction

Loyalty ID	Order Date	Transaction
SARID00060621	8/27/2022	Rp20,000
SARID00095578	2/20/2024	Rp3,485,000
SARID00059562	8/24/2022	Rp250,000
SARID00003072	3/20/2024	Rp2,664,400
SARID00056243	8/20/2022	Rp40,000
SARID00079982	2/11/2023	Rp29,509,400
SARID00034092	7/4/2023	Rp25,000
SARID00083115	10/02/2024	Rp25,702,180
SARID00005721	07/11/2023	Rp83,733,000
SARID00000130	15/08/2024	Rp572,272,148

4.4. Modeling

In this phase, the cleaned dataset will be compared with several segmentation techniques to find the best and most optimal customer segmentation. Furthermore, RFM analysis is used to determine Recency, Frequency, and Monetary value using the R programming language in the RStudio data processing application, and the analysis result will give Sarinah insight related to customers' shopping patterns.

Table 3. Sample data result from RFM analysis

Loyalty ID	R	F	M
SARID00060621	740	1	Rp20,000
SARID00095578	198	1	Rp3,485,000
SARID00059562	743	1	Rp250,000
SARID00003072	169	6	Rp2,664,400
SARID00056243	747	1	Rp40,000
SARID00079982	572	7	Rp29,509,400
SARID00034092	429	1	Rp25,000
SARID00083115	208	8	Rp25,702,180
SARID00005721	303	4	Rp83,733,000
SARID00000130	21	89	Rp572,272,148

In this research, several clustering algorithms, such as K-Means, K-Medoids, Fuzzy C-Means, and DBSCAN, will be compared to find the most effective customer segmentation model based on RFM analysis. After further analysis, K-Means shows the result of a clustering analysis using the *elbow method* with the number of clusters *K* from 1 to 10.

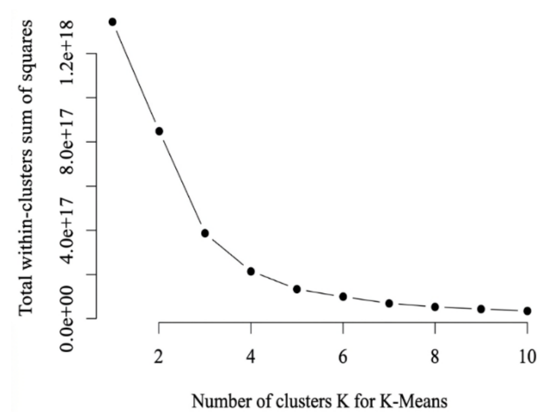

Fig. 4 K-Means clustering

Based on the plot, the values of K from 2 to 7 have been chosen for further analysis with the Davies-Bouldin Index. The reason is that the plot shows an elbow shape, and the elbow method is used to analyze the optimal value of K by identifying the point where the decrease in Within-Cluster Sum of Squares (WCSS) slows down. Hence, analyzing the DBI score for K values within this range will help in identifying the best clusters.

Meanwhile, K-Medoids shows the results of clustering analysis using the silhouette method to find the optimal number of clusters.

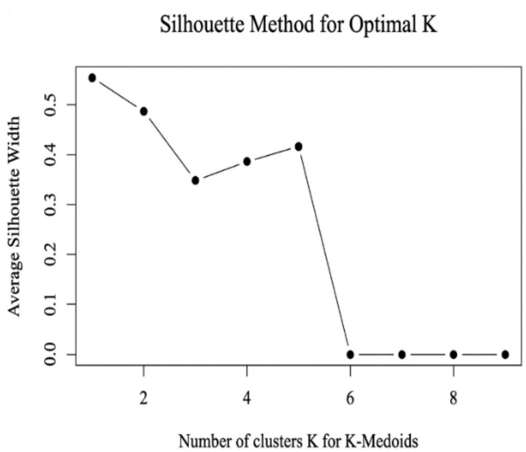

Fig. 5 K-Medoids clustering

From the plot, the optimal *K* using the silhouette method dranges between 2 and 5, and these values will be used for further analysis with DBI. Meanwhile, Fuzzy C-Means shows the results of clustering analysis using *silhouette coefficient* for determining the optimal number of *m*.

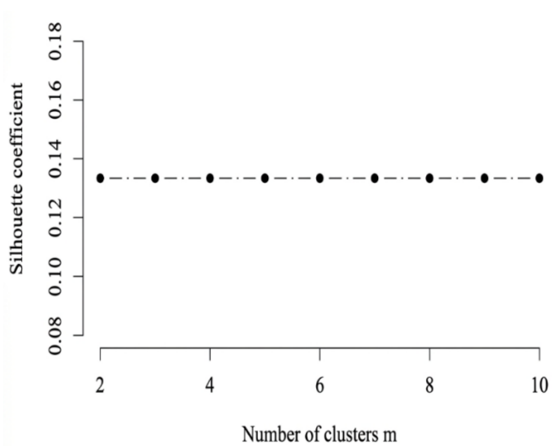

Fig. 6 Fuzzy C-Means clustering

The plot above shows a flat line, which means the silhouette coefficient is consistent in each value of m. The researchers conclude that there is no ideal cluster, and the data might not be suitable for Fuzzy C-Means clustering.

However, this research will proceed to calculate the DBI Score with m is 2. On the other hand, DBSCAN with the silhouette coefficient to find the optimal value of epsilon (ε) shows the result in the plot below.

After further analysis, the plot shows a higher silhouette coefficient on values 6, 7, and 9, which indicates better clusters. Those values will be evaluated with the DBI score to find the best clustering algorithms. After comparing several clustering algorithms, the researcher will evaluate and assess the result analysis with DBI to determine the effectiveness in customer segmentation for Sarinah Department Store. Based on the DBI score, K-Means with cluster value 4 (four) shows the lowest DBI score of 0.571503631365945, which indicates the best performance in Sarinah's customer segmentation.

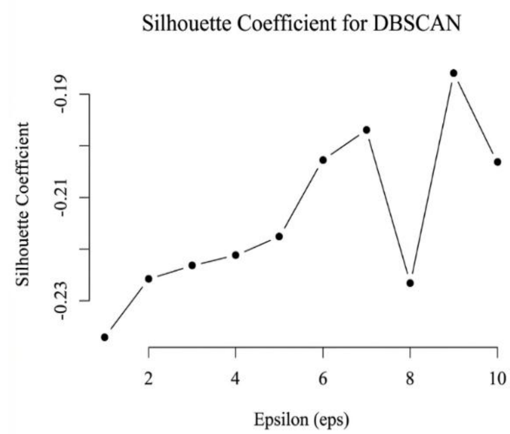

Fig. 7 DBSCAN clustering

Table 4. DBI score

Clusters	Davies-Bouldin Index Score
K-Means	
2	1.16784099600687
3	0.57351674450411
4	0.571503631365945
5	0.579107565790997
6	0,603196680570233
7	0.57769398909698
K-Medoids	
2	1.52435266321267
3	2.04889213077748
4	3.37752704058618
5	5.5933305347463
Fuzzy C-Means	
2	3.730287
DBSCAN	
6	1.124394
7	1.123344
9	1.121166

4.5. Evaluation

Furthermore, the researchers need to calculate the average RFM value with clusters that have the lowest DBI Score to gain more insights for the clustering process.

Table 5. Calculation average RFM values

Cluster	AVG Recency	AVG Frequency	AVG Monetary
1	380	4,15	11.279.202
2	270	9,97	52.811.324
3	614	1,49	917.037
4	17.5	49	514.008.324

After getting the average results of RFM values, the researchers will define threshold values from the calculated average RFM values. These thresholds will help describe customer behavior patterns from each cluster.

By comparing these thresholds, researchers will classify clusters based on high or low levels of recency, frequency, and monetary. This model is supported in previous research, especially in research that uses RFM combined with clustering techniques to find behavior prediction or targeted marketing (Christy et al., 2021; Idowu et al., 2019; Khanfar et al., 2024; Khumaidi et al., 2023).

Table 6. Threshold result

Mean Recency	Mean Frequency	Mean Monetary
320.375	16.15	Rp29.101.971

From the threshold result, Recency shows 320.375, which means the threshold result is 320 days, and the lower values will be preferred as they indicate more recent shopping activity.

Meanwhile, the Frequency shows 16.15, which means the threshold result is 16 transactions, and the higher values will be preferred as they indicate higher shopping frequency. Furthermore, the Monetary shows Rp29.101.971, and the higher values will be preferred as they indicate a higher amount of transactions.

After the thresholds are set, the researchers will categorize the data by labeling each customer and will also determine the customer type based on the RFM score. This process will help segment customers and assist Sarinah in data-driven decision-making.

Table 7. Data labeling

Cluster	AVG Recency	AVG Frequency	AVG Monetary
1	High	Low	Low
2	Low	Low	High
3	High	Low	Low
4	Low	High	High

Table 8. Customer type

Cluster	R	F	M	Customer Type
1	Not Yet	Infrequent	Spend less	New Customers
2	Shop recent	Infrequent	Spend significant	Occasional Customers
3	Not Yet	Infrequent	Spend less	Inactive Customers or at-risk
4	Shop recent	Frequent	Spend significant	Loyal Customers

Table 9. Sample cluster profiling data

Loyalty ID	Cluster	Customer Type
SARID00060621	3	Inactive customers or at-risk
SARID00095578	3	Inactive customers or at-risk
SARID00059562	3	Inactive customers or at-risk
SARID00003072	3	Inactive customers or at-risk
SARID00056243	3	Inactive customers or at-risk
SARID00079982	1	New Customers
SARID00034092	3	Inactive customers or at-risk
SARID00083115	1	New Customers
SARID00005721	2	Occasional customers
SARID00000130	4	Loyal Customers

4.6. Business Interpretation

From the analysis of the cluster profiling that has been conducted, the segmentation of Sarinah Club members shows unique characteristics that require special managerial implications.

Table 10. Segmented customer characteristics count

Cluster	Customer Types	Customer Count
1	New Customers	2.444
2	Occasional Customers	145
3	Inactive Customers or at-risk	30.438
4	Loyal Customers	2

In Cluster 1, the researchers define it as new customers. The data shows 2.444 customers with characteristics of high recency, low frequency, and low monetary value. This is defined as the customer has recently shopped at Sarinah, but has not shopped frequently, and has not made any high-value purchases. To encourage new customers to repeat transactions, Sarinah needs to implement a welcome discount program. Furthermore, Sarinah can implement high-reward referral programs to attract new active members.

Meanwhile, in Cluster 2, the researchers define it as occasional customers. The data show 145 customers with low recency and frequency, but high monetary value. These customers tend to shop infrequently, but their purchases are high-value transactions. Sarinah is supposed to focus on increasing its purchase frequency while maintaining its high transaction value. To encourage occasional customers, Sarinah can offer special discount products based on tier, or Sarinah can provide AI-based recommendations in Sarinah Club application to suggest discounted products based on customer behavior and previous purchases with an accuracy rate of 90%.

In Cluster 3, the researchers define it as inactive customers or at-risk. The data show 30.438 customers with high recency, low frequency, and low monetary value. These customers are at risk of churn and require immediate member reactivation strategies. Sarinah could take immediate action by implementing a time-limited progressive discount program, conducting a customer satisfaction survey with shopping voucher incentives, or holding a flash-sale or a midnight-sale event in the Department Store exclusively for Sarinah Club members to reactivate membership and to encourage them to buy.

Meanwhile, in Cluster 4, the researchers define it as loyal customers. The data show only 2 customers with low recency, high frequency, and high monetary value. Sarinah needs to focus on retaining and expanding these high-value customers by offering an exclusive loyalty program and premium service in the Department Store. Sarinah could also launch a tiered membership program with daily fixed discounts, priority cashier, personal shopper services, and early access to limited and newest collections.

5. Conclusion and Future Work

5.1. Conclusion

This research shows the combination of RFM analysis with several clustering algorithms for Sarinah club members segmentation based on customers' transactions from March 2022 to early September 2025. After comparing four clustering techniques, such as K-Means, K-Medoids, Fuzzy C-Means, and DBSCAN, the analysis shows that K-Means with 4 (four) clusters provides the lowest DBI score, with 0.571503631365945, which indicates an optimal cluster result.

Furthermore, thresholds are calculated within all clusters and used to label the data of each cluster. Final analysis defined Cluster 1 as "New Customers" with 2.444 customers, Cluster 2 as "Occasional Customers" with 145 customers, Cluster 3 as "Inactive Customers or at-risk" with 30.438 customers, and Cluster 4 as "Loyal Customers" with 2 customers. With this analysis and new classification, Sarinah could personalize recommendations and create data-driven managerial strategies. For example, new customers could get a welcome discount program or a reward for the member-get-member program. Occasional customers could get personalized discount product recommendations based on previous purchases. Inactive customers could get a time-limited progressive discount or flash-sale program exclusive to specific categories. Loyal customers might qualify to receive a set amount of discount each day and to be given a personal shopper service. With this research, Sarinah is provided with concrete strategies based on facts to facilitate better decisions, better marketing, and better customer retention.

5.2. Future Work

Future work should consider other approaches to clustering than performed here that could be considered, such as Hierarchical Clustering, Gaussian Mixture Model (GMM), or Self-Organizing Maps (SOM). These methods may uncover behaviors of customers which are not properly captured within the framework of the ongoing research.

Another twist is analyzing churn prediction and Customer Lifetime Value (CLV); by identifying those segments that are at risk of churning out, Sarinah can make efforts to retain those customers. Sentiment Analysis is the other one. Examining feedback, reviews, or even comments on social media from Sarinah Club members can provide greater insight into what people enjoy, what they are frustrated by, and the things in need of improvement in relation to the Sarinah Club services. When combined with transactions, to be more personalized and specific, creating more meaningful and impactful campaigns.

Conflicts of Interest

A statement of authors' Organizational Affiliation disclosing no conflict of interest is included with this publication.

Acknowledgements

Author Contributorship

Frisca Adiesthy Irdiani: Conceptualization, Data Curation, Methodology, Resources, Investigation, Writing Original Draft, Project Administration, and Visualization.

Tuga Mauritsius: Conceptualization, Data Curation, Methodology, Resources, Writing Review and Editing, Supervision, Project Administration, and Validation.

Data Availability

Customer Segmentation Using the K-Means Method with Point of Sales Transaction Sarinah Department Store dataset. The dataset used in this research is available at DOI: https://dx.doi.org/10.21227/scrn-as92.

References

[1] Fitri Andriyani, and Yan Puspitarani, "Performance Comparison of K-Means and DBScan Algorithms for Text Clustering Product Reviews," *Sinkron: Journal and Research in Informatics Engineering*, vol. 6, no. 3, pp. 944-949, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] P. Anitha, and Malini 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, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] A. Joy Christy et al., "RFM Ranking-An Effective Approach to Customer Segmentation," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 10, pp. 1251-1257, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Yulin Deng, and Qianying Gao, "A Study on E-Commerce Customer Segmentation Management based on Improved K-Means Algorithm," *Information Systems and E-Business Management*, vol. 18, no. 4, pp. 497-510, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nur Fitrianti Fahrudin, and Rini Rindiyani, "Comparison of K-Medoids and K-Means Algorithms in Segmenting Customers based on RFM Criteria," *E3S Web of Conferences*, vol. 484, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Lili Fan, "Research on Precision Marketing Strategy of Commercial Consumer Products based on Big Data Mining of Customer Consumption," *Journal of the Institution of Engineers (India): Series C*, vol. 104, no. 1, pp. 163-168, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Hodjat (Hojatollah) Hamidi, and Bahare Haghi, "An Approach based on Data Mining and Genetic Algorithm to Optimizing Time Series Clustering for Efficient Segmentation of Customer Behavior," *Computers in Human Behavior Reports*, vol. 16, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Oluwasurefunmi Idowu et al., "Customer Segmentation based on RFM Model using K-Means, Hierarchical and Fuzzy C-Means Clustering Algorithms," *Fourth International Conference on Informatics and Computing*, pp. 1-6, 2019. [Google Scholar]

[9] Iyad Khanfar, Isra Khanfar, and Mahmoud Odeh, *Customer Segmentation based on RFM Attributes using Machine Learning*, Intelligent Systems, Business, and Innovation Research, Springer, vol. 489, pp. 369-386, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ali Khumaidi et al., "RFM-AR Model for Customer Segmentation using K-Means Algorithm," *E3S Web of Conferences*, Krisnadwipayana University, Jakarta, Indonesia, vol. 465, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Desfiana Suci Rachmahwati, Rachmadita Andreswari, and Faqih Hamami, "Customer Segmentation of Cellular Telecommunication Company using K-Means Algorithm (Case Study of PT Indosat)," *International Conference of Science and Information Technology in Smart Administration 2022*, Denpasar, Bali, Indonesia, pp. 45-50, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Frédéric Ros, Rabia Riad, and Serge Guillaume, "PDBI: A Partitioning Davies-Bouldin Index for Clustering Evaluation," *Neurocomputing*, vol. 528, pp. 178-199, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Karli Eka Setiawan et al., "Clustering Models for Hospitals in Jakarta using Fuzzy C-Means and K-Means," *Procedia Computer Science*, vol. 216, pp. 356-363, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Bei Zhang, Luquan Wang, and Yuanyuan Li, "Precision Marketing Method of E-Commerce Platform based on Clustering Algorithm," *Complexity*, vol. 2021, no. 1, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]