

Systematic Literature Review: Customer Segmentation Analysis in the Retail Industry for Marketing Strategy Optimization

Yanda Rizky Prasetya^{1*}, Firlan Prayoga¹

¹Department of Computer Science
University of Esa Unggul
Jl Arjuna Utara No.9, West Jakarta, Jakarta, Indonesia
yrizky35@gmail.com

Received January 2025; accepted April 2025

ABSTRACT

Customer segmentation is a crucial strategy in the retail industry to enhance marketing effectiveness and understand customer needs. This study adopts a Systematic Literature Review (SLR) approach using the PRISMA framework to analyze research related to customer segmentation from 2019 to 2024. The objective of this study is to evaluate the most frequently used clustering models and association patterns and identify the models that deliver the best performance in supporting marketing strategies. The findings indicate that K-Means is the most commonly used clustering model and excels in generating accurate and efficient customer clusters. Meanwhile, FP-Growth proves to be the best association rule model due to its efficiency in handling large datasets and identifying relevant association rules. This study also identifies several research gaps, including the lack of integration between RFM, K-Means, and FP-Growth models, as well as the limited practical implementation of segmentation and association results in marketing strategies such as up-selling and cross-selling. As a solution, this study proposes an integrated model combining RFM, K-Means, and FP-Growth to provide deeper and more actionable insights in supporting data-driven marketing strategies. This model is expected to enhance customer retention, marketing effectiveness, and operational efficiency in the retail industry.

Keywords: Systematic Literature Review (SLR), PRISMA, K-Means, FP-Growth, RFM

1. Introduction.

In the last few decades, the retail industry has undergone significant changes, driven by technological advancements and changes in consumer behavior in purchasing and obtaining information. To this day, retail companies face intense competition, and therefore they must continue to develop a variety of strategies and innovative ideas to increase the sales revenue generated from product sales. Data analysis within the company has become one of the foundations to find strategies to increase sales and market the company's products. Furthermore, understanding each customer can support the creation of effective communication in offering products by adjusting to their needs and providing personalized services for each customer. However, recognizing the needs of each customer is not an easy task, as customer analysis covers a very broad field. This process may involve various characteristics and complex customer behaviors. Therefore, a strategy is needed to group customers into several groups or clusters. This approach is known as customer segmentation. Customer segmentation aims to group consumers with similar characteristics, needs, and interests into one group, so the company can apply a business model that meets the needs of each segment. This step is taken to make the company's marketing strategy more effective and efficient and to help direct marketing efforts and services according to the preferences

and needs of each segment, thereby increasing customer satisfaction.

This research aims to review and analyze customer segmentation in the retail industry. Clustering, association patterns, or combinations of clustering and association patterns have been widely applied in previous studies for customer segmentation and optimizing marketing strategies. Clustering is a data analysis method that separates a set of data into clusters based on similar characteristics. The goal is to find patterns or structures in data that are not immediately visible. Meanwhile, association patterns are used to identify relationships or correlations between products in a dataset. Additionally, association pattern approaches can generate accurate product recommendations for customers by understanding the purchasing habits and patterns of each customer. This can be utilized by the company as a guide to determine the right marketing strategies, such as cross-selling, upselling, product promotions, and others.

Therefore, the main focus of this systematic literature review is to analyze research on customer segmentation published between 2019-2024 to identify effective clustering and association pattern approaches that are suitable for application in the retail industry. This study is expected to provide comprehensive insights into the implementation of these approaches, which can serve as a reference for retail companies in formulating optimal customer data-based marketing strategies.

2. Methodology

The topic of this research is the analysis of customer segmentation in the retail industry for the optimization of marketing strategies. The following are the Research Questions in this study:

1. What are the most commonly used clustering models for customer segmentation?
2. What association pattern models are used to support marketing strategies?
3. Which clustering model shows the best performance in customer segmentation based on previous research?
4. Which association pattern model shows the best performance in supporting marketing strategies based on previous research?

The method used in this research is Systematic Literature Review (SLR). SLR is a research method and process aimed at identifying and critically evaluating relevant research, while collecting and analyzing data from those studies. As a guideline for conducting SLR, this study uses the PRISMA framework (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). PRISMA is a guideline designed to assist in compiling systematic literature reviews (SLR) and meta-analyses in a complete and transparent manner. PRISMA helps authors filter and select literature. PRISMA has 4 stages, which are:

1. Identification This stage involves identifying several sources or articles relevant to the research topic by searching in databases and filtering out duplicated papers. This study used the Google Scholar database with the assistance of the "Publish or Perish"

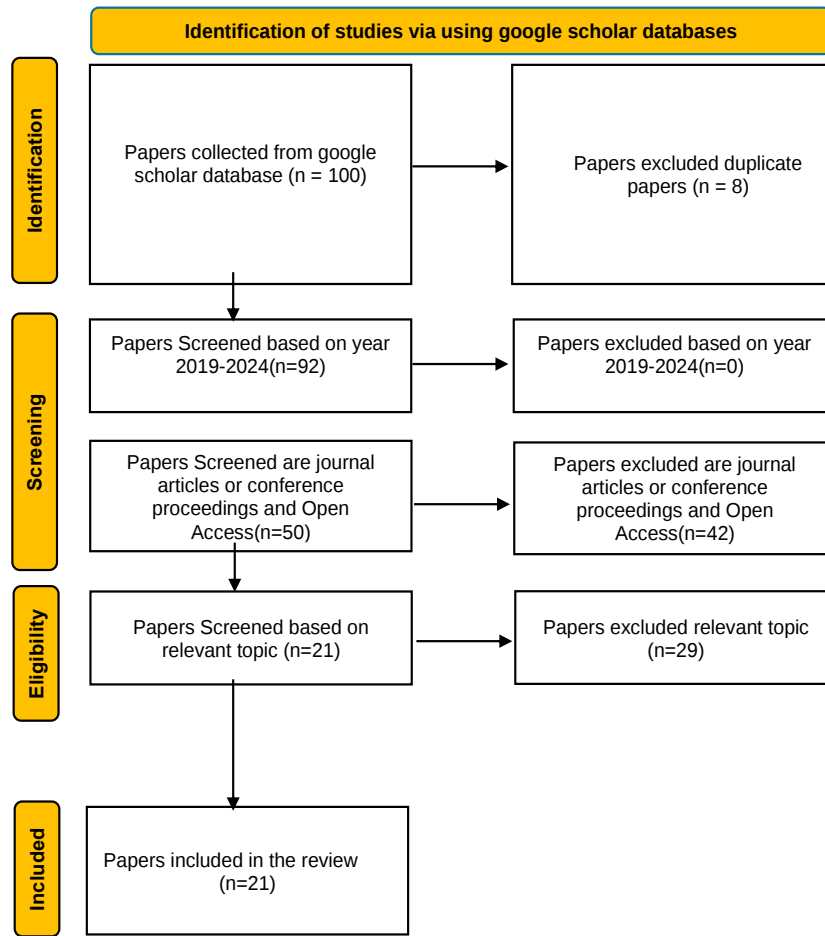
tool to search for journals. "Publish or Perish" is a software tool used by researchers to find relevant references/literature related to the research topic. To search for relevant literature, researchers can perform searches based on keywords and set a publication date range. In this study, the keywords used were "Customer Segmentation," "Clustering," "Association Patterns," and the publication range was from 2019 to 2024.

2. **Screening** This stage involves screening the journals collected from the identification phase. At this stage, the author performs screening based on inclusion and exclusion criteria. Inclusion criteria are used to select articles that meet these criteria, meaning they are deemed suitable for analysis, while exclusion criteria are used to identify articles that are deemed unsuitable. The inclusion and exclusion criteria for this research are as follows:
 - Inclusion Criteria:
 - Articles published between 2019-2024
 - Only journal articles or conference proceedings
 - Research related to the author's research
 - Exclusion Criteria:
 - Articles published outside the range of 2019-2024
 - Research types other than journal articles or conference proceedings
 - Research not related to the author's research.
3. **Eligibility** In this stage, the articles that passed the screening are thoroughly examined for analysis. Elements such as the abstract, title, and full text are used to determine which papers are relevant to the research topic and suitable for further analysis.
4. **Included** This is the final stage of the PRISMA process. Papers/articles that meet the eligibility criteria are included in the final analysis and data synthesis.

3. Results and Discussion

3.1. Results

Figure 1 represents the PRISMA diagram. Based on Figure 1, in the identification stage, a total of 100 articles were collected, 92 articles passed the identification stage, and 8 were excluded due to duplicate papers. Next, 92 articles in the identification stage were screened based on inclusion and exclusion criteria, resulting in 50 articles, with 42 excluded, which were then moved to the eligibility stage. In the eligibility stage, from the remaining 50 articles, further checks were conducted based on the relevance of the topic to the research. In this stage, 29 articles were excluded because they were considered irrelevant to the topic, leaving 21 articles. The 21 articles that met all the inclusion criteria were included in the final analysis for review.



Gambar 1 PRISMA Diagram

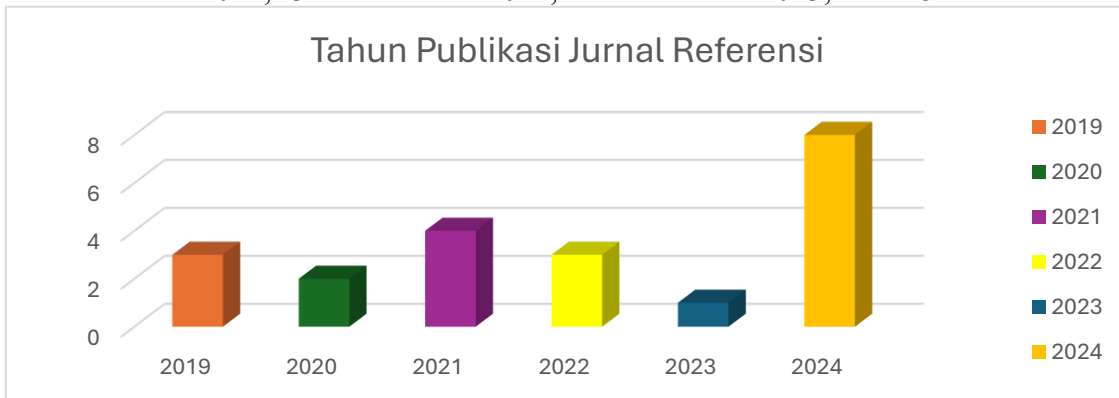
Based on the 21 articles that met the inclusion criteria, the researcher summarized these articles in a table, which can be seen in Table 1.

Table 1: Summary of Selected Articles

Artikel	Pendekatan	Model/Metode
[2]	Pola Asosiasi	Apriori, FP-Growth
[3]	Clustering, Pola Asosiasi	K-Means, FP-Growth
[4]	Clustering	K-Means, metode elbow
[9]	Pola Asosiasi	Apriori, FP-Growth, CRISP-DM
[10]	Pola Asosiasi	FP-Growth
[15]	Clustering	K-Means, CLARA, PAM, Fuzzy C-Means, RFM Analysis
[16]	Clustering	K-Means
[17]	Clustering	K-Means
[18]	Clustering	K-Means, Hierarchical Clustering, Gaussian Mixture Models (GMM)
[19]	Clustering	RFM Model, K-Means
[20]	Clustering	RFM Model, K-Means, Fuzzy C-Means, Repetitive Median K-Means (RM K-Means)
[21]	Clustering	K-Means

[22]	Clustering	RFM Model, K-Means
[23]	Clustering	RFM Model, K-Means, K-Medoids, Fuzzy C-Means, Mini-Batch K-Means
[24]	Clustering	K-Means, DBSCAN
[25]	Pola Asosiasi	Apriori
[26]	Clustering,Pola Asosiasi	K-Means, FP-Growth, Model LRFM
[27]	Clustering	K-Means, K-Medoids, Fuzzy C-Means, LRFM Model
[28]	Pola Asosiasi	Apriori, FP-Growth, CRISP-DM
[29]	Pola Asosiasi	Apriori, FP-Growth, Squeezer
[30]	Clustering,Pola Asosiasi	K-Means, Agglomerative Clustering, FP-Growth

From the summary in Table 1, the researcher created a graphical visualization shown in Figure 2 to display the number of publication years of research articles relevant to the research topic. Based on Figure 2, there were 3 articles published in 2019, 2 articles in 2020, 4 articles in 2021, 3 articles in 2022, 1 article in 2023, and 8 articles in 2024.



Gambar 2 Grafik Tahun Publikasi Jurnal Referensi

3.2 Pembahasan

In this discussion, the researcher will explain and answer the research questions (RQs) that were identified previously. The discussion is as follows:

RQ1: What are the most commonly used clustering models for customer segmentation?

Based on the summary of the selected articles in Table 1, there are several clustering models applied in previous studies. A total of 12 articles used a pure clustering approach, and 3 articles used a combination of clustering and association pattern approaches. Table 2 below shows the articles selected based on the clustering approach and the combination of clustering and association patterns.

Table 2: Articles Selected Based on Clustering Approach

Article	Approach	Model/ Method	Result
[3]	Clustering, Pola Asosiasi	K-Means, FP-Growth	This study uses K-Means to form 5 customer clusters based on transaction data, which is optimal with a Davies-Bouldin Index (DBI) of 0.500.
[4]	Clustering	K-Means, metode elbow	This study applies the K-Means algorithm using the elbow method to determine the optimal number of clusters, which is 4. The clustering is done with variables such as quantity, unit price, and customer ID, resulting in 4 customer clusters with different

			characteristics based on purchase behavior and spending amounts.
[15]	Clustering	K-Means, CLARA, PAM, Fuzzy C-Means, RFM Analysis	This study compares clustering algorithms for customer segmentation in the B2B context, using K-Means, CLARA, PAM, and Fuzzy C-Means. The data used includes POS transaction data from October 2018 to February 2019. K-Means provides the best results in terms of cluster validation and time complexity, compared to Fuzzy C-Means. Based on RFM analysis, three customer clusters are identified: core customers, potential customers, and lost customers, each showing unique RFM characteristics.
[16]	Clustering	K-Means	This study applies the K-Means method and SAPK + K-Means for customer segmentation on purchasing data from an e-commerce platform in Malaysia. The clustering results show three clusters based on customer interest: (1) high interest, (2) moderate interest, and (3) low interest. The data is then used to separate customer segments based on their purchasing behavior in various product categories. Categories with high interest and large profit values tend to focus on equipment and electronics. By segmenting in this way, businesses can focus on the most profitable customers and design appropriate marketing strategies to retain them.
[17]	Clustering	K-Means	This study uses the K-Means algorithm for customer segmentation based on data from Autofurnish.com for 2018 and 2019. Clustering is done considering the Davies-Bouldin Index to determine the optimal number of clusters (k=4 for 2018 and k=8 for 2019). The analysis shows that active user segments mostly come from the 25-34 age group, and Agra city had the most users in 2018. For 2019, user distribution became more spread out with more detailed segmentation, allowing the company to target customers based on demographics and purchasing behavior.
[18]	Clustering	K-Means, Hierarchical Clustering, Gaussian Mixture Models (GMM)	This study compares K-Means, Hierarchical Clustering, and GMM for customer segmentation in the banking sector. The results show that K-Means is effective in clear segmentation with a silhouette score of 0.62, while Hierarchical Clustering provides deeper insights into relationships between segments, though it is less efficient for large datasets. GMM offers the highest flexibility in capturing complex and overlapping customer behaviors but requires significant computational resources.
[19]	Clustering	RFM Model, K-Means	This study uses the RFM model for customer segmentation with the K-Means algorithm on retail transaction data. RFM (Recency, Frequency, Monetary) analyzes customer behavior through purchase frequency, recency of transactions, and monetary value. With k = 3 clusters, the results show that the high-frequency clusters also show high sales values, helping the company identify high-value customers. Silhouette analysis is applied to assess the

			cohesion and separation between clusters, ensuring that the clusters generated are optimal for segmentation based on purchasing patterns.
[20]	Clustering	RFM Model, K-Means, Fuzzy C-Means, Repetitive Median K-Means (RM K-Means)	This study examines the effectiveness of the RFM model in customer segmentation with K-Means, Fuzzy C-Means, and RM K-Means algorithms for online retail transaction datasets. The analysis shows that RM K-Means is more optimal than other methods in terms of execution time and the number of iterations, with a Silhouette Score of 0.49, indicating better cluster separation. This segmentation helps companies identify groups of customers that are more likely to yield higher profits, allowing for more effective and targeted marketing strategies. RFM scoring based on Recency, Frequency, and Monetary values is used to define relevant customer segments according to purchasing behavior.
[21]	Clustering	K-Means	This study uses the K-Means algorithm for customer segmentation based on annual income and expenditure scores. By optimizing the cluster values using the Elbow, Silhouette, and Gap Statistic methods, it was found that 6 clusters are optimal for data segmentation. These clusters include various customer categories, from those with high to low income and spending, enabling the company to target services and products more effectively according to the needs of each segment.
[22]	Clustering	RFM Model, K-Means	This study uses the RFM model and the K-Means algorithm to analyze customer purchasing behavior in online transaction data from Beijing, China. Through PCA analysis to determine the weight of each RFM indicator, customers are classified into four groups based on frequency, purchase amount, and last purchase time. The results show an increase of 529 active customers, a 279% increase in total purchase volume, and a 101.97% increase in total consumption.
[23]	Clustering	RFM Model, K-Means, K-Medoids, Fuzzy C-Means, Mini-Batch K-Means	This study uses the RFM model combined with the additional attribute Discount Proportion (DP) for customer segmentation on e-commerce data from the MurahJaya888 store on the Shopee platform. By applying four clustering algorithms, it was found that Mini-Batch K-Means provided the highest Silhouette Score (0.5002), and K-Means had the best CH Index (1056). The segmentation results divided customers into four groups (Platinum, Gold, Silver, Bronze), which showed different responses to discounts, allowing for more effective marketing strategies and increased customer loyalty.
[24]	Clustering	K-Means, DBSCAN	This study compares the K-Means and DBSCAN algorithms for customer segmentation in e-commerce, considering variables such as customer ID, gender, age, city, membership type, total spending, items purchased, average rating, discounts, and satisfaction. The results show that DBSCAN gives a higher Silhouette Score (0.680) compared to K-Means

			(0.546), indicating better cluster separation, while the DBSCAN Davies-Bouldin index is higher, indicating less compact clusters. K-Means is suitable for segmentation requiring balanced clusters, while DBSCAN is effective for identifying niche markets and unique customer behaviors in datasets with varying densities.
[26]	Clustering,Pola Asosiasi	K-Means, FP-Growth, LRFM Model	This study implements the K-Means algorithm for customer segmentation in the Shopee marketplace using the LRFM (Length, Recency, Frequency, Monetary) model and generates three customer clusters: "Consistent Customers," "Active Customers," and "Superactive Customers".
[27]	Clustering	K-Means, K-Medoids, Fuzzy C-Means, LRFM Model	This study uses a combination of the K-Means algorithm, Agglomerative Clustering, and FP-Growth for product recommendations in frozen food sales. The results show that the dataset clustered using K-Means generated the most association rules, with 18 rules, 11 of which had a lift ratio above 1, indicating the effectiveness of product segmentation for recommendations.

Based on Table 2, it can be concluded that the K-Means model is the most frequently used model in this study, followed by K-Medoids and Fuzzy C-Means.

RQ2: What are the association pattern models used to support marketing strategies?

Based on the summary of the selected articles in Table 1, there are several models in the association pattern approach applied in previous studies. A total of 6 articles used a pure association pattern approach, and 3 articles used a combination of association patterns and clustering. Table 3 below shows the articles selected based on the clustering approach and the combination of clustering and association patterns.

Tabel 3 Artikel yang dipilih berdasarkan pendekatan *clustering*

Article	Approach	Model/ Method	Result
[2]	Pola Asosiasi	Apriori, FP-Growth	In this study, FP-Growth is used to find association patterns in each cluster. It was found that 8 products are frequently purchased together.
[3]	Clustering, Pola Asosiasi	K-Means, FP-Growth	In this study, FP-Growth is useful for the association process. Based on the Association Rule formed in each cluster, it is known that there are 8 items/products.
[9]	Pola Asosiasi	Apriori,FP-Growth, CRISP-DM	This study aims to compare the Apriori and FP-Growth algorithms to identify association patterns in sales transaction data at Maestro Jakarta Cafe & Space. Using data from November 2023 to April 2024, the results show that FP-Growth is superior in processing speed compared to Apriori and requires less memory. Both algorithms generated 22 identical association rules.

[10]	Pola Asosiasi	FP-Growth	This study applies FP-Growth to analyze product association patterns at ABC Business Center to formulate up-selling and cross-selling strategies. Transaction data from July 2023 to December 2023 show that rules with a minimum support of 5% and confidence of 70% can be applied for cross-selling strategies, while rules with minimum support of 5% and confidence of 10% are suitable for up-selling. Recommendations generated include product bundling and offering higher-value products to boost sales.
[25]	Pola Asosiasi	Apriori	This study uses the Apriori algorithm to analyze transaction data in a minimarket to build a sales strategy based on product purchase patterns. The analysis results found several significant association rules, such as the frequent purchase of instant noodles followed by eggs (support 33.33%, confidence 76.92%). All the discovered rules have a lift ratio > 1, indicating strong validity and potential for applying sales strategies.
[26]	Clustering,Pola Asosiasi	K-Means, FP-Growth, Model LRFM	Clustering, Association Patterns - K-Means, FP-Growth, LRFM Model FP-Growth is used to identify product association rules that are frequently purchased together.
[28]	Pola Asosiasi	Apriori, FP-Growth, CRISP-DM	Association Patterns - Apriori, FP-Growth, CRISP-DM This study compares the Apriori and FP-Growth algorithms to determine recommended items in promotional packages. Using transaction data from Dewimedialestari, it was found that FP-Growth is more efficient in processing time compared to Apriori. Association results with a minimum support of 15% and confidence of 20% produce promotional rules such as "If buying TAS and PENA, then also buying BLOCKNOTE" with 37% confidence.
[29]	Pola Asosiasi	Apriori, FP-Growth, Squeezer	Association Patterns - Apriori, FP-Growth, Squeezer This study analyzes customer purchase patterns at Sinar Pagi Hardware Store using Apriori, FP-Growth, and Squeezer algorithms. The FP-Growth algorithm generated 52 association rules with a total rule strength of 1.16, while Apriori generated 194 rules with the same strength. Squeezer generated 7 clusters based on product similarity, helping store owners understand customer preferences.
[30]	Clustering,Pola Asosiasi	K-Means, Agglomerative Clustering, FP-Growth	Clustering, Association Patterns - K-Means, Agglomerative Clustering, FP-Growth This study uses a combination of K-Means, Agglomerative Clustering, and FP-Growth algorithms for product recommendations in frozen food sales. The results show that the dataset clustered using K-Means generated the most association rules, with 18 rules, 11 of which had a lift ratio above 1, indicating the effectiveness of product segmentation for recommendations.

Based on Table 3, the most frequently used models are Apriori and FP-Growth, both of which support marketing strategies by identifying products that are frequently purchased together and supporting the development of strategies such as bundling or product recommendations.

RQ3: Which clustering model shows the best performance in customer segmentation based on previous studies?

Based on several analyses of the articles from previous studies in Table 1, it can be concluded that K-Means is the most widely used clustering model in research and shows superior performance in customer segmentation. This model excels in terms of efficiency and generates clear and consistent clusters. This is supported by strong statements from previous research studies. According to research [4], the implementation of the K-Means model for customer segmentation provides easily understandable clusters and supports more targeted marketing strategies. Research [15] compared clustering models for customer segmentation in the B2B context using K-Means, CLARA, PAM, and Fuzzy C-Means. The results show that K-Means provided the best results in terms of cluster validation and time complexity compared to Fuzzy C-Means. In research [18], a comparison of clustering models for customer segmentation in the banking sector was conducted using K-Means, Hierarchical Clustering, and GMM. The results showed that K-Means is effective in producing clear segmentation with a silhouette score of 0.62. Research [27] compared the algorithms K-Means, K-Medoids, and Fuzzy C-Means for customer segmentation in the cosmetics industry. The K-Means algorithm provided the best results with the lowest Davies-Bouldin Index (DBI) of 0.167 with six clusters, compared to K-Medoids (DBI = 0.192) and Fuzzy C-Means (DBI = 0.502). K-Means was found to be the most effective for customer segmentation based on loyalty and transaction frequency characteristics.

Based on the results of previous studies, it can be concluded that the K-Means model is chosen as the best clustering model for customer segmentation due to its ability to provide efficient, accurate clusters that meet business needs across various industries, including the retail industry.

RQ4: Which association pattern model shows the best performance in supporting marketing strategies based on previous research?

Based on several analyses of the articles from previous studies in Table 1, the association pattern model that shows the best performance in supporting marketing strategies is the FP-Growth model. FP-Growth excels at finding association rules in large datasets more efficiently than Apriori, thereby speeding up the process and reducing memory consumption. This is supported by the results from previous studies. In research [2], it was found that FP-Growth is 0.5 seconds faster than Apriori, which takes 0.6 seconds. In another study [9], the aim was to compare the Apriori and FP-Growth algorithms for identifying association patterns in sales transaction data at Maestro Jakarta

Cafe & Space, and the results showed that FP-Growth outperformed Apriori in processing speed and required less memory. Another study [28] compared the Apriori and FP-Growth algorithms to determine items recommended in promotional packages and found that FP-Growth is more efficient in processing time than Apriori. Based on the results of previous studies, it can be concluded that the FP-Growth model is the best association pattern model for supporting marketing strategies due to its efficiency in processing large data and its ability to find relevant association rules.

3.3 Discussion

In the 21 articles that were analyzed and reviewed, the researcher identified gaps in each article, and these gaps need to be addressed so that they can become a reference for improvements or developments in future research. Table 4 presents the gaps identified in the selected articles:

Tabel 4 Analisa GAP Terhadap artikel yang dipilih

Article	Model/Method	GAP Analysis
[2]	Apriori, FP-Growth	This study does not assess how the association rules found through the Apriori and FP-Growth algorithms can be applied in practical marketing strategies, such as bundling promotions or product placement. Additionally, there is no analysis regarding the comparison of algorithm performance on larger datasets to determine the efficiency of the algorithm in different conditions.
[3]	K-Means, FP-Growth	Customer segmentation based on RFM is not equipped with an evaluation of the direct impact on marketing strategies or customer loyalty. Additionally, the variation of threshold parameters in the RFM model that could affect segmentation results is not discussed.
[4]	K-Means, metode elbow	The study is limited to RFM-based segmentation without evaluating the long-term effectiveness of marketing strategies for each segment. The inclusion of demographic data such as age or customer location was also not considered to improve the segmentation accuracy.
[9]	Apriori, FP-Growth, CRISP-DM	This study compares Apriori and FP-Growth for association rules in the cafe sector, but does not implement these rules in concrete marketing strategies. The impact on sales or customer satisfaction has not been analyzed, which could serve as a useful evaluation step.
[10]	FP-Growth	This study only uses FP-Growth for up-selling and cross-selling strategies without evaluating whether these rules successfully increase sales. Parameter variations such as support and confidence values were also not explored, which could affect the relevance of product recommendations.
[15]	K-Means, CLARA, PAM, Fuzzy C-Means, RFM Analysis	This study focuses on RFM-based segmentation with centroid clustering algorithms but does not include additional variables that could affect customer segmentation in a B2B context, such as business category and geographical location. Furthermore, the study does not explore the effectiveness of segmentation in implementing specific marketing strategies.
[16]	K-Means	This study uses e-commerce purchase behavior data with centroid-based clustering algorithms but does not explore additional variables such as demographics or more specific product preferences. Moreover, the direct impact of this segmentation on

		marketing strategies and long-term customer loyalty has not been analyzed.
[17]	K-Means	This study does not review additional variables that could enrich customer behavior analysis, such as product preferences or specific categories of interest. Furthermore, no further analysis is conducted on how this segmentation could be integrated into effective marketing strategies to enhance customer retention and conversion.
[18]	K-Means, Hierarchical Clustering, Gaussian Mixture Models (GMM)	This study is limited to segment analysis without evaluating the direct impact of segmentation on marketing strategies or customer loyalty. Moreover, it does not explore advanced machine learning algorithms such as deep learning, which may provide more precise segmentation for datasets with complex characteristics.
[19]	RFM Model, K-Means	This study focuses solely on RFM-based segmentation without evaluating the direct impact of this segmentation on marketing strategies or customer loyalty. Additionally, the study does not discuss parameter variations that might affect segmentation results, such as the RFM threshold values for different customer categories.
[20]	RFM Model, K-Means, Fuzzy C-Means, Repetitive Median K-Means (RM K-Means)	This study is limited to RFM-based segmentation analysis without exploring the long-term effectiveness of the recommended marketing strategies for each customer segment. Moreover, additional variables such as customer demographic data are not considered, which could potentially improve segmentation accuracy and the relevance of marketing strategies.
[21]	K-Means	This study is limited to income and expenditure score-based analysis without considering demographic variables or product preferences that could improve segmentation accuracy. Moreover, the impact of this segmentation on marketing strategies and customer retention is not analyzed in depth.
[22]	RFM Model, K-Means	This study focuses on customer segmentation without exploring how variations in CRM strategies could be applied across different industries or analyzing the direct effect of segmentation on long-term customer loyalty. Moreover, it does not consider the use of demographic variables or additional characteristics that could enrich segmentation results.
[23]	RFM Model, K-Means, K-Medoids, Fuzzy C-Means, Mini-Batch K-Means	This study focuses on adding the Discount Proportion attribute in RFM but has not measured the direct impact of discount strategies on customer retention and loyalty over the long term. Furthermore, it does not explore the use of demographic data or other behavioral factors that could enrich segmentation and marketing strategy effectiveness.
[24]	K-Means, DBSCAN	This study does not explore the combination of K-Means and DBSCAN methods as a hybrid approach to improve segmentation effectiveness. Moreover, the impact of segmentation on marketing strategies or customer loyalty has not been deeply analyzed, so the segmentation results have not been directly linked to practical business implementation.
[25]	Apriori	This study only uses the Apriori algorithm without comparing it to other algorithms that might be more efficient in handling large datasets. Additionally, it has not assessed the impact of implementing the resulting association rules on increasing sales or customer satisfaction directly.
[26]	K-Means, FP-Growth, Model LRFM	This study focuses on product segmentation and association using historical purchase data but has not assessed the impact of applying

		the segmentation results or association rules on future shopping behavior. Additionally, further analysis related to purchase preferences based on specific times or seasons outside of Ramadan has not been explored, which could provide insights for long-term sales strategies.
[27]	K-Means, K-Medoids, Fuzzy C-Means, LRFM Model	This study has not explored the direct impact of segmentation results on marketing strategies or personalized offerings to customers. Moreover, it has not considered other variables such as demographic or psychographic data that could enrich customer segmentation and result in more targeted strategies based on individual preferences.
[28]	Apriori, FP-Growth, CRISP-DM	This study only compares the time efficiency of the algorithms without evaluating the impact of the promotional recommendations generated on sales effectiveness or customer satisfaction. Additionally, the study has not tested the validity of the association rules beyond historical data to see if the same patterns remain relevant during different periods.
[29]	Apriori, FP-Growth, Squeezer	This study compares the effectiveness of algorithms without evaluating the practical implementation of the association rules or clustering generated in optimizing sales or stock management. Additionally, this study does not consider seasonal factors or time trends, which could affect customer purchasing patterns in hardware stores.
[30]	K-Means, Agglomerative Clustering, FP-Growth	This study does not directly evaluate how the implementation of product recommendations based on association results can increase sales or stock management. Moreover, the study has not tested the effectiveness of other clustering methods that may be better suited for products with similar characteristics.

Based on the gaps presented in Table 4, the researcher has proposed a model for future research, which integrates RFM, K-Means, and FP-Growth to provide a more structured and innovative approach to customer segmentation and purchasing pattern analysis. Unlike previous studies that used a combination of K-Means and association patterns, this proposed model offers several key advantages:

- This proposed model utilizes RFM to gain an in-depth understanding of customer behavior before performing segmentation with K-Means.
- In previous studies, association patterns were often applied universally without considering customer segmentation. This proposed model addresses this weakness by analyzing purchasing patterns within each cluster, namely royal customers, middle customers, and lost customers.
- This proposed model explicitly connects the results of the analysis with practical marketing strategies, such as up-selling for middle customers and cross-selling for royal customers. This has not been found in previous research, which only stopped at data analysis without implementing specific marketing strategies.

4. Conclusions

This study has conducted a systematic literature review of 21 articles focusing on customer segmentation in the retail industry. The analysis results show that K-Means is the most widely used clustering model and has proven effective in generating efficient and accurate customer clusters. Additionally, FP-Growth emerged as the best association pattern model due to its

efficiency in processing and its ability to handle large datasets and find relevant association rules. However, there are still several gaps that have not been addressed in previous studies. As a solution to these gaps, this study proposes an integrated model combining RFM, K-Means, and FP-Growth to produce more specific customer segmentation. This model divides customers into three main clusters: royal customers, middle customers, and lost customers, with specific purchasing patterns analyzed within each cluster. The results of this analysis provide strategic insights that can be used to support loyalty programs, up-selling strategies, and targeted promotions based on the characteristics of each customer cluster. By implementing this model, companies are expected to improve marketing effectiveness, strengthen customer retention, and maximize sales opportunities. The proposed model not only offers significant practical contributions but also addresses the gaps in previous research, making it a relevant solution to the challenges in customer analysis within the retail industry.

REFERENCES

- [1] D. Grewal, A. L. Roggeveen, and J. Nordfält, “The Future of Retailing,” *Journal of Retailing*, vol. 93, no. 1, pp. 1–6, Mar. 2017, doi: 10.1016/j.jretai.2016.12.008.
- [2] Islamiyah, P. L. Ginting, N. Dengen, and M. Taruk, “Comparison of Priori and FP-Growth Algorithms in Determining Association Rules,” in *International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, IEEE, 2019. doi: 10.1109/ICEEIE47180.2019.8981438.
- [3] S. Genjang Setyorini, K. Sari, L. Rahma Elita, and S. A. Putri, “Analisis Keranjang Pasar Menggunakan Algoritma K-Means dan FP-Growth pada PT. Citra Mustika Pandawa,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 41–46, Apr. 2021.
- [4] E. Febrianty, L. Awalina, and W. I. Rahayu, “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail,” *Jurnal Teknologi dan Informasi (JATI)*, vol. 13, 2023, doi: 10.34010/jati.v13i2.
- [5] A. Wicaksono, F. A. Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Fuzzy C-Means Clustering berdasarkan RFM Model pada E-Commerce (Studi Kasus: E-Commerce XYZ),” 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] S. Wahyuni, S. Informasi, F. Ilmu Komputer, U. Mulia Jalan Pahlawan No, A. Kota Samarinda, and P. Kalimantan Timur, “SEGMENTASI PELANGGAN BERDASARKAN ANALISIS RECENCY, FREQUENCY, MONETARY MENGGUNAKAN ALGORITMA K-MEANS,” *Jurnal SimanteC*, vol. 12, no. 1, Dec. 2023.
- [7] S. Bahri, D. Marisa Midyanti, and P. Korespondensi, “Penerapan Metode K-Medoids Untuk Pengelompokan Mahasiswa Berpotensi Drop Out,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.
- [8] M. Fathurrahman, A. Rizky Pratama, T. Al-Mudzakir, and U. Buana Perjuangan, “Perbandingan Algoritma Apriori Dan Fp Growth Terhadap Market Basket Analysis Pada Data Penjualan Bakery,” *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 4, no. 2, pp. 266–274, 2023.
- [9] M. Raihan and Sutisna, “Analisis Perbandingan Algoritma Apriori dan FP-Growth untuk Menentukan Strategi Penjualan Pada Maestro Jakarta Cafe & Space,” 2024. [Online]. Available: <https://journal.stmiki.ac.id>
- [10] Nabiihah, M. Afdal, R. Novita, and Mustakim, “Implementasi Association Rule Untuk Rekomendasi Strategi Up-Selling dan Cross-Selling Produk Menggunakan FP-Growth,” *Indonesian Journal of Computer Science*, vol. 13, no. 3, 2024.
- [11] F. Achmad, O. Nurdian, and Y. A. Wijaya, “Analisa Pola Transaksi Pembelian Konsumen Pada Toko Ritel Kesehatan Menggunakan Algoritma Fp-Growth,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, 2023, [Online]. Available: www.researchgate.net

- [12] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J Bus Res*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.
- [13] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," Mar. 29, 2021, *BMJ Publishing Group*. doi: 10.1136/bmj.n71.
- [14] A. Farzana Sukaryo and R. Mulyati Mustika Sari, "SYSTEMATIC LITERATURE REVIEW : KEMAMPUAN NUMERASI SISWA DALAM PEMBELAJARAN MATEMATIKA Corresponding Author *," *Jurnal Theorems (The Original Reasearch Of Mathematics)*, vol. 8, no. 2, 2024.
- [15] N. R. Maulina, I. Surjandari, and A. M. M. Rus, "Data Mining Approach for Customer Segmentation in B2B Settings using Centroid-Based Clustering," *Institute of Electrical and Electronics Engineers(IEEE)*, 2019.
- [16] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
- [17] R. Punhani, V. P. S. Arora, A. Sai Sabitha, and V. K. Shukla, "Segmenting e-commerce customer through data mining techniques," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1742-6596/1714/1/012026.
- [18] M. Abu Sufian Mozumder *et al.*, "Optimizing Customer Segmentation in the Banking Sector: A Comparative Analysis of Machine Learning Algorithms," *Journal of Computer Science and Technology Studies*, Aug. 2024, doi: 10.32996/jcsts.2024.6.4.1.
- [19] 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.
- [20] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, Dec. 2021, doi: 10.1016/j.jksuci.2018.09.004.
- [21] E. Y. L. Nandapala and K. P. N. Jayasena, "The practical approach in Customers segmentation by using the K-Means Algorithm," in *2020 IEEE 15th International Conference on Industrial and Information Systems, ICIIS 2020 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Feb. 2020, pp. 344–349. doi: 10.1109/ICIIS51140.2020.9342639.
- [22] J. Wu *et al.*, "Retracted: An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm," *Math Probl Eng*, vol. 2023, Aug. 2023, doi: 10.1155/2023/9873736.
- [23] V. H. Antonius and D. Fitrianah, "Enhancing Customer Segmentation Insights by using RFM + Discount Proportion Model with Clustering Algorithms," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, 2024, [Online]. Available: www.ijacsa.thesai.org
- [24] A. Suryaputra Paramita and T. Hariguna, "Comparison of K-Means and DBSCAN Algorithms for Customer Segmentation in E-commerce," *Journal of Digital Market and Digital Currency*, vol. 1, no. 1, pp. 43–62, Jun. 2024.
- [25] R. Andika Johan, R. Himilda, and N. Auliza, "Penerapan Metode Association Rule Untuk Strategi Penjualan Menggunakan Algoritma Apriori," *Jurnal Teknologi Informatika (J-TIFA)*, vol. 2, no. 2, pp. 1–7, Sep. 2019.
- [26] M. Arif Saifudin, H. Endah Wahanani, and A. Junaidi, "IMPLEMENTASI ALGORITMA ASOSIASI FP-GROWTH DAN KLASIFIKASI K-MEANS TERHADAP POLA PEMBELIAN KONSUMEN DI MARKETPLACE SHOPEE," 2024.
- [27] H. Syukron, M. Fauzi Fayyad, F. Junita Fauzan, Y. Ikhsani, and U. Rizkya Gurning, "Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 76–83, Oct. 2022.
- [28] M. Mariko and U. Amikom Yogyakarta, "Perbandingan Algoritma Apriori Dan Algoritma Fp-Growth Untuk Rekomendasi Item Paket Pada Konten Promosi," *EXPLORE*, vol. 11, no. 2, 2021.
- [29] Faris Syaifulloh, Eva Yulia Puspaningrum, and M. Muharram Al Haromainy, "Analisis Pola Pembelian Pelanggan Menggunakan Algoritma Squeezer, Apriori dan FP-Growth Pada Toko Bangunan," *Modem : Jurnal Informatika dan Sains Teknologi*, vol. 2, no. 3, pp. 134–147, Jul. 2024, doi: 10.62951/modem.v2i3.153.

- [30] R. Najmil Huda, R. Fitriadi, and A. Wibowo, “OPTIMIZATION PRODUCT RECOMMENDATION USING K-MEANS, AGGLOMERATIVE CLUSTERING AND FP-GROWTH ALGORITHM,” *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 4, pp. 953–960, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.1901.