Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

Integration of RFM and Apriori Algorithm in a CRM System for Targeted Promotion

Salman Alfarisi[1*], Ari Purno Wahyu Wibowo[2]
[1,2]Universitas Widyatama, Indonesia
[1*]salman.alfarisi@widyatama.ac.id, [2]ari.purno@widyatama.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 13-11-2026
Accepted: 17-11-2026
Published: 23-11-2026

Keywords:
Apriori Algorithm; Association Rules; Customer Relationship Management; Customer Segmentation; RFM Analysis; Targeted Promotion.

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

The growth of small and medium-scale multi-vendor e-commerce platforms in Indonesia has increased the need for intelligent customer relationship management (CRM), yet most platforms managed by a single administrator still apply uniform promotions without considering differences in customer purchase behavior, which reduces marketing effectiveness and customer loyalty. This study aims to design and implement an integrated CRM system that combines Recency, Frequency, Monetary (RFM) analysis for customer segmentation with the Apriori algorithm for purchase pattern analysis to generate targeted promotion recommendations. The system was developed using Node.js, Python, React.js, and MongoDB, and encapsulates the entire analytical process from data preprocessing to recommendation output. The algorithms were validated using the Online Retail II dataset, which yielded 779,425 clean transaction records from 5,878 customers after preprocessing. RFM values were scored using a quintile-based approach and mapped into seven behavioral segments, after which the Apriori algorithm was applied specifically to the Champions segment with minimum support of 0.02, confidence of 0.30, and lift of 1.0. Champions, the highest-value segment, comprised 1,570 of these customers and accounted for 63.35 per cent of all orders, and rule mining on that segment generated 41 strong association rules with the highest lift reaching 26.41. The implemented system displayed these segments and the resulting product recommendations through an administrator dashboard. Transaction records from an Indonesian MSME retailer were further assessed, indicating that catalog breadth and repeat purchase behavior are preconditions for applying the approach in that setting.

INTRODUCTION

Indonesia's e-commerce transaction value grew consistently from IDR 106 trillion in 2018 to IDR 401 trillion in 2021 (Hendra et al., 2024), a trajectory that has expanded the space for micro, small, and medium enterprises to operate their own multi-vendor e-commerce platforms alongside established marketplaces. Such platforms are generally operated with limited resources and administered by a single operator who manages all vendors within one system. This limitation prevents them from developing customer analytics capabilities comparable to those of large marketplaces, which are supported by dedicated data science teams and sophisticated recommendation engines.

Without customer analytics, administrators of such platforms design promotions on intuition rather than transaction evidence, and campaigns are distributed uniformly to all customers regardless of how recently they purchased, how frequently they transact, or how much they contribute in revenue. Sudrajat et al. (2025) reported the same condition in an Indonesian fashion retailer, where limited insight into purchasing behavior and the absence of personalized marketing were identified as the causes of customer attrition. High-value customers therefore receive no differentiated treatment, customers who begin to disengage are not identified before they are lost, and promotional budgets are spent on segments unlikely to respond. Customer relationship management (CRM) is the established framework for this condition, since it organizes customer data into differentiated engagement strategies rather than uniform outreach. The systems operating on these platforms, however, act mainly as transactional record keepers, storing orders and customer details without converting them into segment-level insight.

Addressing this problem requires knowing both which customers to target and what to offer them, and previous studies have demonstrated the effectiveness of each approach individually. RFM analysis has proven reliable for customer segmentation across a range of transaction-based settings (Devi et al., 2024; Irawan et al., 2025; Susilowati et al., 2022), while the Apriori algorithm effectively discovers co-purchase patterns within transaction data (Hendra et al., 2024; Kaban et al., 2024). Several studies have attempted to combine both. Hoque et al. (2024) demonstrated that applying the Apriori algorithm to high-value segments produces stronger and more actionable rules than analyzing the entire transaction dataset, Pai et al. (2026) combined the two techniques to separate high-lift customers from the remainder of an e-commerce customer base, while Yıldız et al. (2023) generated cluster-based association rules for a fashion retailer. Such integration nevertheless remains uncommon, and a further separation persists between analysis and practice. Studies that do combine both approaches, including Sibarani et al. (2024) in an Indonesian café setting,

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

present them as analytical experiments rather than as systems an operator can run, while systems that implement CRM with RFM for promotional purposes, such as Sudrajat et al. (2025), stop at a segmentation dashboard and do not extend to the purchase patterns within the segments they produce. What remains unaddressed is an operational CRM system in which both techniques are embedded and usable by an administrator without data science expertise.

To address this gap, this study designs and implements a CRM system that integrates RFM analysis and the Apriori algorithm to support targeted promotions in a multi-vendor e-commerce platform. The contributions are threefold. The first is the integration of RFM segmentation and Apriori association-rule mining into an operational CRM system for multi-vendor platforms. The second is the application of the Apriori algorithm exclusively to the high-value Champions segment, following the approach proposed by Hoque et al. (2024). The third is an administrator dashboard that translates analytical results into actionable targeted promotion strategies without requiring data science expertise. The analytical modules are validated on a benchmark retail dataset, while transaction records from an Indonesian MSME retailer are analyzed to assess the conditions under which the approach can be applied in that setting.

LITERATURE REVIEW

Customer relationship management refers to the strategy and supporting systems through which a business acquires, retains, and strengthens its relationships with customers by treating them according to their value rather than uniformly (Sudrajat et al., 2025). Its analytical dimension depends on the ability to convert stored transaction records into knowledge about customer behavior. Reviewing customer segmentation for targeting in e-commerce, Alves Gomes and Meisen (2023) found that segmentation is the mechanism most frequently used for this purpose, and that a substantial share of the applications they surveyed define segments through thresholds on behavioral values rather than through clustering, most of them using RFM. Systems operating in smaller businesses, however, rarely reach this analytical dimension at all, holding transaction records without deriving any distinction between the customers who generated them.

Among segmentation techniques, the RFM model remains one of the most widely adopted because it relies on three observable behavioral dimensions that can be derived directly from transaction records. Devi et al. (2024) applied it to segment institutional customers of a pharmaceutical distributor and used the resulting scores to prioritize marketing effort, while Susilowati et al. (2022) combined it with clustering to direct promotional activity toward the most responsive target groups, and Irawan et al. (2025) confirmed its continuing relevance in more recent retail settings. What these studies share is that segmentation identifies which customers deserve attention but does not indicate what should be offered to them.

Association rule mining addresses the complementary question. The Apriori algorithm discovers products that are frequently purchased together and evaluates them through support, confidence, and lift, and it has been applied to promotional decisions in retail contexts by Kaban et al. (2024). Hunyadi et al. (2025) compared it with FP-Growth on e-commerce transaction data and reported that although FP-Growth is faster at large scale, Apriori remains appropriate where the transaction volume is moderate. These studies typically apply the algorithm to an entire transaction dataset, so the resulting rules describe the behavior of the customer population as a whole and cannot indicate for whom a particular product combination is most relevant.

Combining the two techniques is therefore a natural response. Hoque et al. (2024) showed that restricting Apriori to high-value segments yields rules with stronger confidence and lift than analyzing the full dataset, Pai et al. (2026) mined rules separately for customer groups distinguished by value, and Yıldız et al. (2023) produced cluster-based rules for a fashion retailer. Chen and Gunawan (2023) extended the same pairing by adding periodicity and a customer engagement index to the RFM representation before clustering and rule mining. Sibarani et al. (2024) applied it to an Indonesian café, grouping customers by RFM before mining menu combinations. These contributions nevertheless remain analytical experiments conducted outside any operating environment, leaving open the question of how such an integration can be embedded within a CRM system that a single non-technical administrator is able to run on a multi-vendor platform.

METHOD

This study applies a quantitative research approach to develop a CRM system that integrates RFM analysis and the Apriori algorithm for targeted promotion on a multi-vendor e-commerce platform. The two techniques are applied sequentially rather than as separate procedures, since association rule mining is performed only on the segment produced by the RFM analysis, and the results are presented through an administrator dashboard that translates them into promotion strategies without requiring data science expertise. The Online Retail II dataset is used to validate both analytical modules, while transaction records from an Indonesian MSME fashion retailer are analyzed to assess the applicability of the approach in a small-scale retail setting. The research process is carried out through eight stages, as illustrated in Figure 1.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

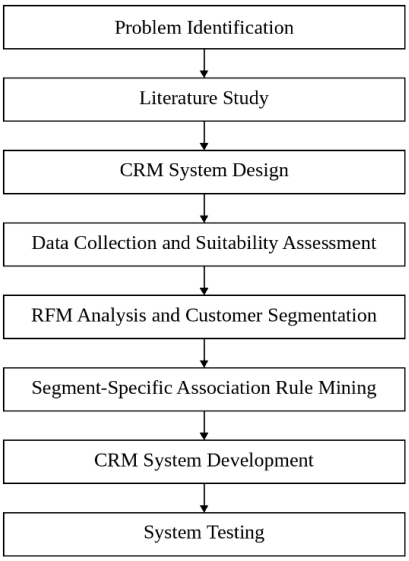


Figure 1. Research Stages

Problem Identification

The problem identification stage was carried out by examining how promotion is currently handled on a small multi-vendor e-commerce platform. The platform stores every order together with the identity of the customer who placed it, but it provides no feature for ranking customers by value and no feature for revealing which products tend to be purchased together. Promotion is consequently handled as a single broadcast addressed to the entire customer list, and any attempt to differentiate it would require the administrator to export the transaction table and process it manually outside the system. This stage established the requirement for analytical functions embedded in the system itself rather than performed separately.

Literature Study

A literature study was conducted to determine which analytical techniques were appropriate for the condition identified in the previous stage. RFM was selected for segmentation rather than demographic or survey-based approaches because it requires only attributes already contained in the order records of the platform, so no additional data collection is needed before the system can operate. Apriori was selected for purchase pattern analysis because its output takes the form of explicit rules with interpretable metrics rather than a model whose reasoning is opaque to the administrator. The decision to restrict rule mining to the highest-value segment follows the approach applied by Hoque et al. (2024) and Pai et al. (2026).

CRM System Design

The system was designed as a single operational environment consisting of three modules and an administrator dashboard. The segmentation module reads the stored transaction records, performs preprocessing and RFM scoring, and writes the resulting customer segments back into the database. The rule mining module retrieves the customers assigned to the highest-value segment, extracts their transactions, and stores the association rules obtained from them. The promotion module maps each segment and each rule to a promotion that the administrator can execute. The two analytical modules operate sequentially rather than in parallel, because the rule mining module takes the customer identifiers produced by the segmentation module as its input filter. The relationship between the modules, the database, and the dashboard is presented in Figure 2.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

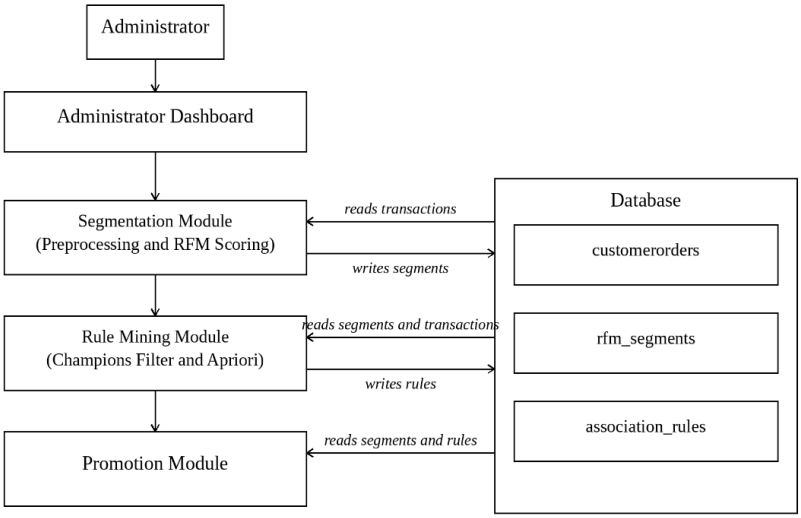


Figure 2. System Architecture

Figure 2 arranges the system in two parts. On the left, the administrator reaches the three analytical modules through the dashboard, and the vertical arrows give the order in which the modules run. On the right, the database holds three collections, and the labelled arrows give what each module reads and writes: the segmentation module reads customerorders and writes rfm_segments, the rule mining module reads both because it filters the transactions by segment membership and writes association_rules, and the promotion module reads the segments and the rules. Results travel between the modules through the database rather than directly, so a module can be re-executed on its own.

The system involves one actor, namely the platform administrator, whose core functions are running the RFM analysis, generating association rules from the resulting high-value segment, and creating promotions targeted at a selected segment. These are supported by authentication, customer data management, and the interfaces through which the segments and the product recommendations derived from the rules are displayed. The platform also provides vendor accounts through which sellers manage their own product catalog and store information, but these functions are operational rather than analytical and fall outside the scope of this study, so they are not included in the diagram. The interaction between the actor and the administrator functions is presented in Figure 3.

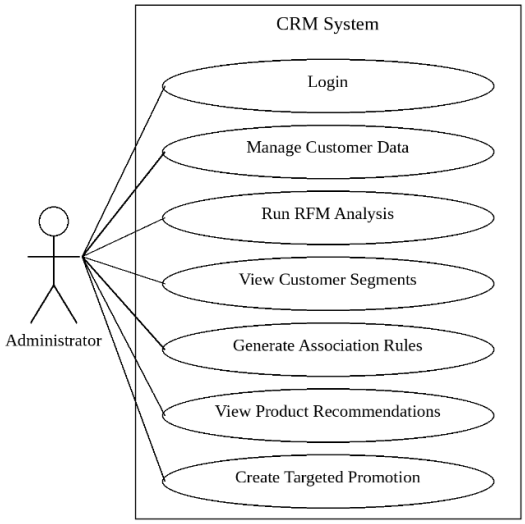


Figure 3. Use Case Diagram

Figure 3 shows the seven use cases available to the administrator, who is the only actor, inside the boundary of the CRM system. Two of them, Login and Manage Customer Data, are supporting functions, and the remaining five make up the analytical workflow. Run RFM Analysis produces the segments, View Customer Segments displays them, Generate Association Rules mines the transactions of the selected segment, View Product Recommendations presents the products derived from the rules, and Create Targeted Promotion directs an offer at that segment. The diagram draws no ordering between them, but the last three depend on the segments written by Run RFM Analysis.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

Data Collection and Suitability Assessment

Two datasets were collected for two different purposes. The first is Online Retail II, a public benchmark dataset that records the transactions of a United Kingdom based online retailer between 2009 and 2011. This dataset is used to validate the analytical modules, because its transaction volume, observation period, and multi-item basket structure allow both quintile-based RFM scoring and association rule mining to be executed under standard parameters. The second dataset consists of order export files obtained from an Indonesian MSME fashion retailer operating on a marketplace platform, covering the period from January to November 2024. This dataset is used to assess the conditions under which the proposed approach can be applied in a small multi-vendor setting. The attributes and characteristics of both datasets are summarized in Table 1.

Preprocessing was applied to both datasets before analysis. For the benchmark dataset, returned invoices, records without a customer identifier, records with a non-positive quantity or unit price, and duplicate rows were removed in that order. Product counts were derived from the product description field after leading and trailing spaces were removed, and the same field was used as the item identifier in the association rule mining. For the MSME dataset, only orders with a completed status were retained, monetary values were corrected for the thousand separator convention used in the export file, and duplicate order-level rows were removed before aggregation at the customer level, because the payment total is repeated on every item row belonging to the same order.

The suitability of the MSME dataset was then assessed against two criteria derived from the operating requirements of the two techniques. The first criterion is the dispersion of purchase frequency, because quintile scoring requires the frequency distribution to produce more than two distinct score levels. The second criterion is basket composition, measured by the average number of items per basket, the proportion of single-item baskets, and the highest support obtained by any two-itemset relative to the minimum support threshold applied in this study. The outcome of this assessment is reported in the Result section.

Table 1. Dataset Description

Aspect	Benchmark Dataset	MSME Dataset
Source	Online Retail II, public repository	Marketplace order export files
Domain	United Kingdom online retailer	Indonesian MSME fashion retailer
Observation period	December 2009 to December 2011	January 2024 to November 2024
Raw records	1,067,371 transaction records	4,394 completed orders
Attributes used	Invoice, Customer ID, Invoice Date, Description, Quantity, Unit Price	Order Number, Buyer Username, Order Date, Product Name, Total Payment, Order Status
Role in this study	Validation of the RFM and Apriori modules and seeding of the system	Suitability assessment of the proposed approach

RFM Analysis and Customer Segmentation

The preprocessed transactions were aggregated per customer to obtain the three RFM dimensions. Recency was calculated as the distance in days between a reference date and the most recent transaction of a customer, as formulated in (1).

$$R_i = t_{ref} - t_{i,last} \quad (1)$$

where R_i is the recency of customer i, t_{ref} is the reference date defined as one day after the latest transaction in the dataset, and $t_{i,last}$ is the date of the last transaction of customer i. Frequency was calculated as the number of distinct orders placed by a customer during the observation period, as formulated in (2).

$$F_i = n(T_i) \quad (2)$$

where T_i is the set of distinct orders belonging to customer i. Monetary was calculated as the total value of those orders, as formulated in (3).

$$M_i = \Sigma_{j \in Ti} (q_j \times p_j) \quad (3)$$

where q_j and p_j are the quantity and the unit price of the items contained in order j. Each dimension was then converted into a score ranging from one to five using quintile boundaries. Recency was scored inversely, so that the most recent customers received the highest score, while frequency and monetary were scored directly. The frequency score and the monetary score were combined into a single value, as formulated in (4).

$$fmScore_i = round((SF_i + SM_i) / 2) \quad (4)$$

where SFi and SMi are the quintile scores of frequency and monetary for customer i. Quintile boundaries were obtained by ranking customers on each dimension and dividing the ranked values into five groups of equal size, with ties resolved by customer identifier so that the assignment is reproducible across executions, and the combined score was rounded half-up. The reference date was taken as the day after the latest transaction in the dataset rather than the date of execution, so that recency reflects the observation period of the data. The pair formed by the recency score and the combined score positions each customer in a five by five grid, in which every cell is mapped to one of seven segments, as presented in Table 2.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

Table 2. RFM Segment Grid

R Score	fmScore 1	fmScore 2	fmScore 3	fmScore 4	fmScore 5
5	Recent	Recent	Loyal	Champions	Champions
4	Recent	Promising	Loyal	Champions	Champions
3	Promising	Promising	Loyal	Loyal	At Risk
2	Hibernating	Hibernating	At Risk	At Risk	At Risk
1	Lost	Lost	Hibernating	At Risk	At Risk

Table 2 maps each pair of scores to a segment. The recency score is read down the rows and the combined frequency and monetary score across the columns, so the upper right corner holds customers who purchased recently and transact both often and in high value, and the lower left holds those who did neither. Champions occupies the four cells in which both scores are four or five, Recent and Promising cover newer customers whose value is not yet established, Loyal runs down the middle of the grid, and Hibernating and Lost occupy the lower left. At Risk is placed wherever a high combined score meets a falling recency score, including the cell where recency is three and the combined score is five, because those customers carry the largest revenue exposure. Each of the twenty-five cells maps to exactly one of seven segments.

Segmentation was performed using this rule-based grid rather than a clustering algorithm. Cluster boundaries produced by algorithms such as K-Means shift whenever the model is recomputed on updated transaction data, and the resulting clusters carry no fixed business meaning, so they must be reinterpreted before they can be used operationally. The quintile grid is deterministic, produces the same segment labels on every execution, and can be read directly by an administrator, which makes it suitable for a system that recalculates segments as new transactions arrive. A non-clustering treatment of RFM within a CRM context has also been applied by Sudrajat et al. (2025).

Segment-Specific Association Rule Mining

Association rule mining was applied exclusively to the customers assigned to the Champions segment, following the approach applied by Hoque et al. (2024) and Pai et al. (2026), so that the resulting rules describe the purchasing behavior of the customers to whom the promotion will be directed. The orders belonging to these customers were retrieved from the transaction collection and transformed into a basket format in which each row represents one order and each column one product.

The Apriori algorithm generates frequent itemsets iteratively, forming candidate k-itemsets from the frequent (k − 1)-itemsets and discarding any candidate whose subset is not frequent. The support of an itemset is the proportion of transactions that contain it, as formulated in (5).

$$\text{Support}(X) = \sigma(X) / N \tag{5}$$

where $\sigma(X)$ is the number of transactions containing itemset X and N is the total number of transactions in the segment. Rules were then formed from the frequent itemsets and evaluated using confidence, as formulated in (6), and lift, as formulated in (7).

$$\text{Confidence}(X \rightarrow Y) = \text{Support}(X \cup Y) / \text{Support}(X) \tag{6}$$

$$\text{Lift}(X \rightarrow Y) = \text{Confidence}(X \rightarrow Y) / \text{Support}(Y) \tag{7}$$

where X is the antecedent and Y is the consequent of the rule. Confidence expresses the conditional probability that a transaction containing the antecedent also contains the consequent, while lift expresses how much more often the two occur together than would be expected if they were independent. A minimum support of 0.02, a minimum confidence of 0.30, and a lift greater than one were used as thresholds. The threshold was set at this level because lower values produced several hundred rules on the same segment, a volume an administrator without data science expertise cannot act upon, while higher values removed the strongest product pairs. Verda and Muselli (2025) report the same sensitivity, noting that the number of rules grows sharply as the threshold is relaxed. Only rules satisfying all three thresholds were stored in the database and displayed as product recommendations on the dashboard. The consequents of the stored rules were collected and deduplicated by product name, each product retaining the highest lift it attained in any rule, and the five products with the highest lift were presented as the recommendations attached to the segment.

CRM System Development

The system was developed as a web application with a separate administrator interface. The server side was built using Node.js and the Express framework, which hosts the preprocessing and RFM scoring module and exposes the analytical results through application programming interfaces. The Apriori procedure was implemented in Python using the mlxtend library and is invoked by the server as a subprocess, so that the mining routine can be executed on demand without a separate manual step. MongoDB was used as the database, storing raw customer orders, segmentation results, and extracted association rules in separate collections. The client side was built using React, which renders the customer segments, the association rules, and the promotion recommendations. Integration between the modules follows the

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

sequential design shown in Figure 2, which keeps both techniques inside a single operational environment.

System Testing

System testing was conducted using the black-box method to verify that each function operates according to its specification, covering data import, execution of the RFM analysis, display of the customer segments, generation of the association rules, display of the product recommendations, and creation of a targeted promotion. In addition, the analytical output was verified by requiring every retained rule to satisfy a lift value greater than one, which confirms that the reported associations are positive rather than the result of chance co-occurrence.

RESULT

Dataset Suitability Assessment

The MSME dataset contained 4,394 completed orders placed by 4,197 unique customers across 36 products between 1 January and 30 November 2024, with a median order value of IDR 42,000 and a mean of IDR 53,177. The first criterion was not satisfied. A total of 4,027 customers, or 95.95 per cent of the customer base, recorded a frequency of one, while 170 customers, or 4.05 per cent, recorded a frequency of two or more, and the highest frequency observed was five. Quintile scoring on the frequency dimension therefore produced only two of the five score levels, so the customers could not be distributed across the segmentation grid.

The second criterion was likewise not satisfied. The mean basket contained 1.204 items and 84.98 per cent of the baskets contained a single item, while the highest support obtained by any two-itemset was 0.0091. No item pair therefore reached the minimum support threshold of 0.02 applied in this study, nor a more permissive threshold of 0.01, and no association rule was generated. Both analytical modules consequently could not be executed on this dataset. The remaining results reported in this section were therefore obtained from the benchmark dataset.

Preprocessing Results

The benchmark dataset contained 1,067,371 transaction records before preprocessing. Returned invoices, records without a customer identifier, records with a non-positive quantity or price, and duplicate rows were removed, which reduced the dataset to 779,425 clean transaction records, or 73.02 per cent of the original volume. The cleaned dataset covered 36,969 orders placed by 5,878 unique customers across 5,241 unique products, and these records formed the input for the RFM analysis. The preprocessing results are presented in Table 3.

Table 3. Preprocessing Results

Step	Records
Raw transaction records	1,067,371
Removed: returned invoices	19,494
Removed: no customer identifier	242,257
Removed: non-positive quantity or price	71
Removed: duplicate rows	26,124
Clean transaction records (73.02%)	779,425
Orders	36,969
Unique customers	5,878
Unique products	5,241

RFM Segmentation Results

The RFM analysis was applied to the 5,878 customers obtained from preprocessing, and every customer was assigned to one of the seven segments defined by the segmentation grid. The distribution of customers across the segments is presented in Table 4.

Table 4. RFM Segmentation Results

Segment	Number of Customers	Percentage
Champions	1,570	26.71%
Loyal	1,032	17.56%
Promising	481	8.18%
Recent	197	3.35%
At Risk	1,006	17.11%
Hibernating	763	12.98%
Lost	829	14.10%
Total	**5,878**	**100%**

Champions formed the largest segment with 1,570 customers, or 26.71 per cent of the customer base, with an average recency of 20.87 days, an average frequency of 14.92 orders, and an average monetary value of £7,914.09 during the observation period. These customers accounted for 23,419 of the 36,969 orders in the dataset, or 63.35 per

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

cent of all orders.

Together with Loyal, the two highest-value segments accounted for 2,602 customers, or 44.27 per cent of the base. The three segments indicating reduced or discontinued engagement, namely At Risk with 1,006 customers, Lost with 829 customers, and Hibernating with 763 customers, accounted for 2,598 customers, or 44.20 per cent. Promising and Recent, which contain customers with shorter purchase histories, accounted for 678 customers, or 11.53 per cent.

Association Rules

Association rule mining was applied to the 23,419 orders placed by the 1,570 customers assigned to the Champions segment, of which 21,491 contained two or more items. At the minimum support threshold of 0.02, the Apriori algorithm generated 223 frequent itemsets.

Rules were then formed from these itemsets and filtered using the minimum confidence threshold of 0.30 and a lift greater than one, and 41 rules satisfied all three thresholds. The five rules with the highest lift are presented in Table 5.

Table 5. Association Rules

No.	Antecedent	Consequent	Support	Confidence	Lift
1	Green Regency Teacup and Saucer	Roses Regency Teacup and Saucer	0.021	0.806	26.41
2	Roses Regency Teacup and Saucer	Green Regency Teacup and Saucer	0.021	0.687	26.41
3	Dolly Girl Lunch Box	Spaceboy Lunch Box	0.021	0.674	18.92
4	Spaceboy Lunch Box	Dolly Girl Lunch Box	0.021	0.590	18.92
5	Alarm Clock Bakelike Green	Alarm Clock Bakelike Red	0.024	0.686	18.06

The strongest association linked the Green Regency Teacup and Saucer with the Roses Regency Teacup and Saucer, with a support of 0.021, a confidence of 0.806, and a lift of 26.41. The confidence indicates that 80.6 per cent of the orders containing the antecedent also contained the consequent, and the lift that the two occurred together about 26 times more often than independence would predict. Rules appear in both directions because lift is symmetric, while confidence differs by antecedent: a customer who bought the green variant was considerably more likely to also buy the roses variant than the reverse.

All 41 rules obtained a lift greater than one, satisfying the criterion defined for the analytical output in the testing stage, and were written to the database as product recommendations attached to the Champions segment.

System Implementation

The system is demonstrated using the benchmark dataset so that the interface output can be compared directly with the tables above. The segmentation interface is shown in Figure 4. The administrator executes the RFM analysis from this page, and the system displays the seven segments together with the number of customers assigned to each of them, corresponding to the distribution reported in Table 4. Selecting a segment opens the list of customers belonging to it.

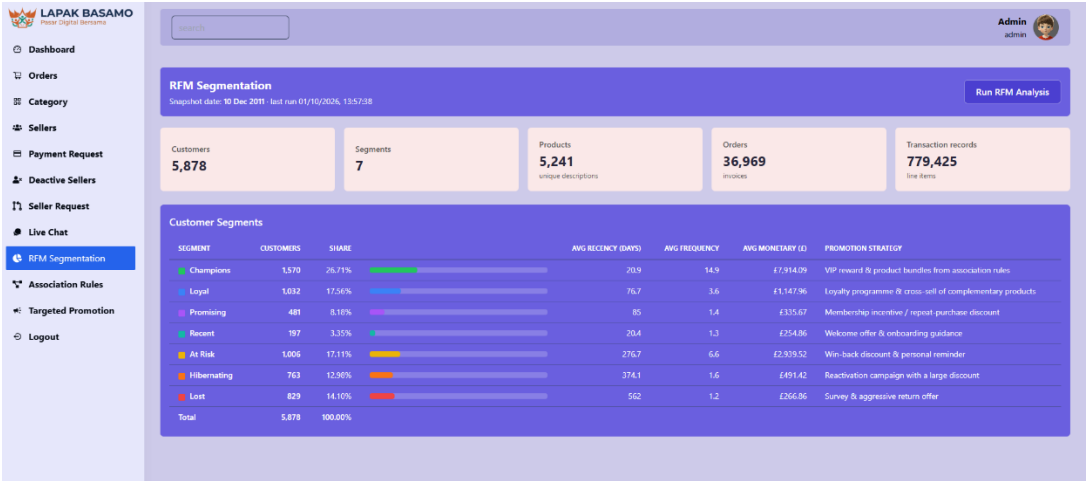


Figure 4. Customer Segmentation Interface

The association rule interface is shown in Figure 5. The rules are generated from the Champions segment using the thresholds defined in the method, and each rule is displayed with its antecedent, consequent, support, confidence, and lift, corresponding to the values reported in Table 5. The five products with the highest lift among the rule

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 6, Number 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

consequents are presented to the administrator as recommendations for the segment.

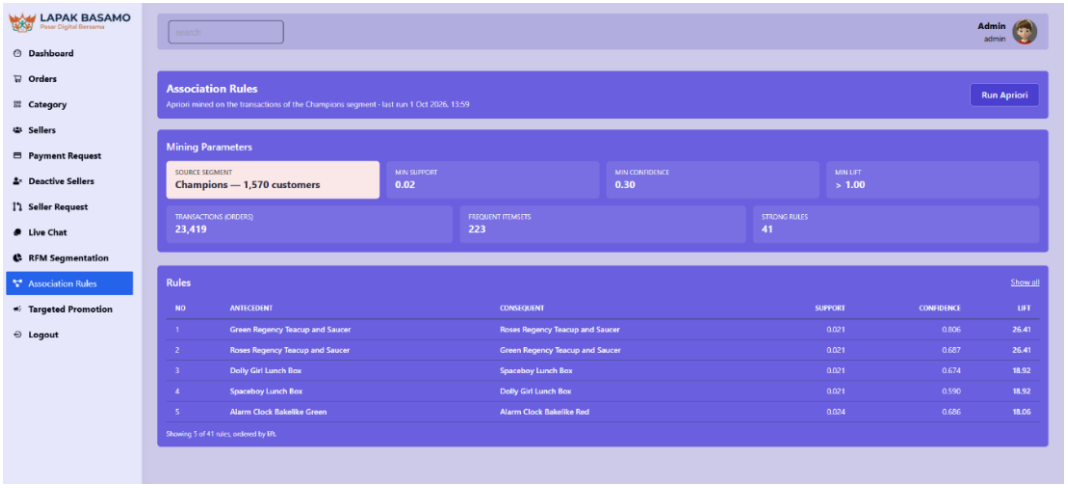


Figure 5. Association Rule Interface

The promotion module attaches a strategy to each segment according to its transactional profile, and for the Champions segment the recommended products are drawn from the association rules rather than selected manually. The mapping between the segments and the promotion strategies is presented in Table 6.

Table 6. Segment-to-Promotion Strategy Mapping

Segment	Transactional Profile	Promotion Strategy
Champions	Recent, frequent, high value	VIP reward and product bundles from association rules
Loyal	Consistent purchases, moderate to high value	Loyalty program and cross-sell of complementary products
Promising	Recent, low frequency	Membership incentive and repeat-purchase discount
Recent	Very recent, short history	Welcome offer and onboarding guidance
At Risk	Declining recency, high past value	Win-back discount and personal reminder
Hibernating	Long inactivity, low value	Reactivation campaign with a large discount
Lost	No recent activity, lowest value	Survey and aggressive return offer

The targeted promotion interface is shown in Figure 6. The administrator selects a segment, reviews the recommended products, and composes a promotion that is directed only at the customers belonging to that segment.

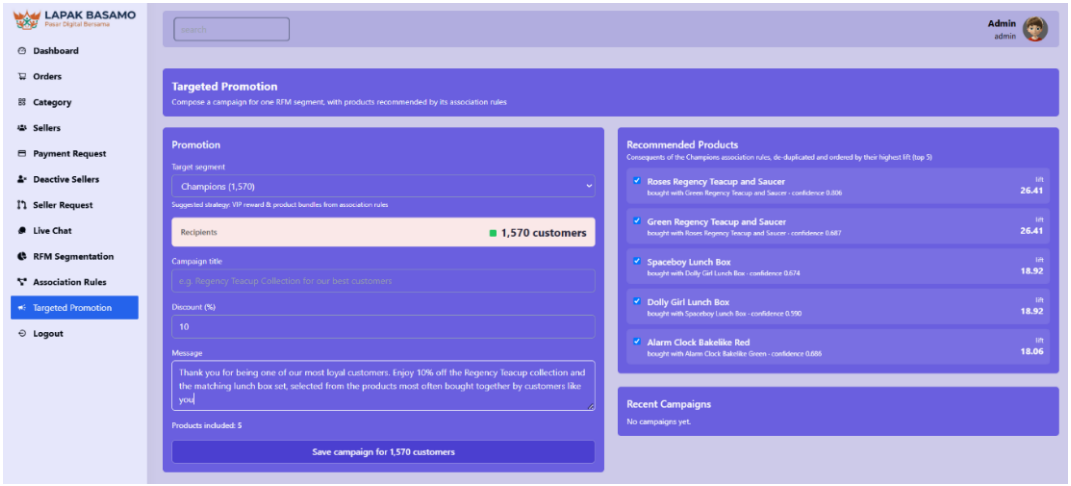


Figure 6. Targeted Promotion Interface

Black-box testing was carried out on the seven functions defined in the testing stage. Each function produced the output expected for the input given, and the segment names, customer counts, and rule metrics displayed by the interface matched the values obtained from the analytical modules.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

DISCUSSION

The segmentation also produced a concentration of value that matters for how the promotion module is used. Champions accounts for 26.71 per cent of the customers but 63.35 per cent of all orders, so a campaign addressed to a quarter of the customer base reaches most of the purchasing activity.

The effect of restricting rule mining to one segment was examined by repeating the analysis on the entire customer base under identical thresholds. The Champions segment yielded 41 rules against 15 from the full base, while the median lift was 8.04 against 10.69 and the highest lift 26.41 against 27.88. The segment therefore produced a larger number of qualifying patterns but not stronger associations, which does not reproduce for this dataset the improvement in rule strength reported by Hoque et al. (2024). The value of the restriction in this study lies elsewhere. The rules describe the purchasing behavior of the customers who will actually receive the promotion, so the products attached to a Champions campaign are derived from what Champions themselves buy together rather than from the behavior of the customer population as a whole, among which the Champions are a minority of 26.71 per cent.

A second difference concerns how the segments themselves are produced. Hoque et al. (2024) and Yıldız et al. (2023) both applied K-Means to the RFM values before mining rules, and Chen and Gunawan (2023) extended the representation with periodicity and a customer engagement index before clustering. These approaches obtain segments that fit the particular dataset closely, but the cluster boundaries are recomputed whenever the model is executed again, and the resulting groups must be interpreted before a promotion can be attached to them. In a system that recalculates segments as new transactions arrive and is operated by an administrator without data science expertise, a deterministic grid that returns the same seven labels on every execution is more practical, since each label can carry a fixed promotion strategy.

The outcome of the suitability assessment reflects the product structure of the MSME retailer rather than a deficiency in data volume. The catalog consists of 36 largely substitutable items within a single category, so customers select one variant rather than combining complementary products, which leaves little item co-occurrence for association rule mining to detect. The same structure explains the absence of repeat purchase that prevented the frequency dimension from forming five quintile levels. The benchmark dataset presents the opposite structure, with 5,241 products across many categories, and the strongest rule obtained from it associates two variants of the same tableware range, which are complementary within a set rather than alternatives to each other. This contrast suggests that catalog breadth is a precondition for market basket analysis, and that the technique is not universally applicable to small retailers however much transaction data they accumulate.

The results also indicate that the value of these techniques depends on where they are executed. Đukanović et al. (2022) concluded that association rule results become useful for direct marketing only once they are integrated into a comprehensive CRM system, and the studies discussed above stop at the analytical stage, presenting rules and segments as findings rather than as functions an operator can run. Embedding both techniques in a single environment, where the segmentation output is persisted and consumed directly by the rule mining module, allows the analysis to be repeated on current data and converted into a promotion without any step performed outside the system.

Several limitations should be noted. The analytical modules were validated on a dataset drawn from a United Kingdom online retailer whose customer base includes a substantial proportion of wholesale buyers, so transactions per customer are far more numerous than in Indonesian small-scale retail, and the validation holds for data of that character rather than for MSME data in general. The suitability assessment used records from a single retailer in one product category, so its outcome describes that retailer rather than Indonesian MSMEs as a whole. The system was evaluated through functional testing and the consistency between interface output and analytical results, but it was not operated with active vendors and live customers, so the effect of the promotions on customer response was not measured. The mapping between segments and promotion strategies was defined from the transactional profile of each segment and has not been validated against the judgement of practitioners.

CONCLUSION

This study designed and implemented a CRM system that integrates RFM analysis and the Apriori algorithm to support targeted promotion on a multi-vendor e-commerce platform. The two techniques operate sequentially within a single environment, where the segmentation output is written to the database and consumed by the rule mining module, so that rules are extracted only from the highest-value segment. Applied to the benchmark dataset, the segmentation assigned 5,878 customers to seven segments, of which 1,570 were classified as Champions and accounted for 63.35 per cent of all orders, and rule mining on that segment produced 223 frequent itemsets and 41 strong rules, all with a lift greater than one. The administrator dashboard presents these results and allows a promotion to be composed for a selected segment without any analytical step performed outside the system.

The system was validated on a benchmark retail dataset, while the assessment of Indonesian MSME transaction data indicates that catalog breadth and repeat purchase behavior are preconditions for applying the approach in that setting. Further work should deploy the system on a platform with active vendors in order to measure the effect of the generated promotions on customer response, and should examine whether a broader catalog formed by aggregating several vendors on one platform produces the item co-occurrence that a single small retailer does not.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 6, **Number** 4, November 2026
https://doi.org/10.47709/brilliance.v6i4.10032

REFERENCES

Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management, 21*(3), 527–570. https://doi.org/10.1007/s10257-023-00640-4

Chen, A. H.-L., & Gunawan, S. (2023). Enhancing retail transactions: A data-driven recommendation using modified RFM analysis and association rules mining. *Applied Sciences, 13*(18), 10057. https://doi.org/10.3390/app131810057

Devi, R. F., Siswanto, F. H., Azkia, N., & Heikal, J. (2024). Customer segmentation based on RFM analysis as the basis of marketing strategy: Case study of PT Pertiwi Agung Pharmaceutical Industry (Landson). *BUDGETING: Journal of Business, Management and Accounting, 5*(2), 766–774. https://doi.org/10.31539/budgeting.v5i2.8994

Đukanović, M., Rogić, S., Novićević, L., Popović-Bugarin, V., & Jovanović, M. (2022). Application of Apriori algorithm for CRM improvement: Case study from Montenegro. In *Proceedings of the 2022 8th International Conference on Computer Technology Applications* (pp. 48–56). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3543712.3543733

Hendra, H., Hermawan, A., & Edy, E. (2024). Smart product recommendations in web e-commerce: Leveraging Apriori algorithm for market basket analysis. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 18*(3).https://doi.org/10.22146/ijccs.89075

Hoque, M. J., Islam, M. S., & Mohtasim, S. A. (2024). Optimizing decision-making through customer-centric market basket analysis. *Journal of Operational and Strategic Analytics, 2*(2), 72–83. https://doi.org/10.56578/josa020201

Hunyadi, I. D., Constantinescu, N., & Ţicleanu, O.-A. (2025). Efficient discovery of association rules in e-commerce: Comparing candidate generation and pattern growth techniques. *Applied Sciences, 15*(10), 5498. https://doi.org/10.3390/app15105498

Irawan, F., Amaluddin, F., & Wijayanti, A. (2025). E-commerce customer segmentation based on RFM analysis using DBSCAN algorithm to improve marketing strategy. *Jurnal EMT KITA, 9*(4), 1850–1859. https://doi.org/10.35870/emt.v9i4.5375

Kaban, E., Darmawiguna, I. G. M., & Kesiman, M. W. A. (2024). Optimizing customer purchase insights: Apriori algorithm for effective product bundle recommendations. *Brilliance: Research of Artificial Intelligence, 4*(2), 747–756. https://doi.org/10.47709/brilliance.v4i2.4981

Pai, P.-Y., Lin, S.-W., & Lu, W.-M. (2026). Integration of association rule mining and RFM analysis with machine learning for e-commerce customer value segmentation: A sustainable retail perspective. Quality & Quantity, 60(1), 87–125. https://doi.org/10.1007/s11135-025-02259-8

Sibarani, J. M. P., Akbar, Y., Sutisna, & Setiawan, K. (2024). Implementation of RFM analysis to enhance sales patterns of food and beverages at Bonjour Café and Resto using the Apriori algorithm. *International Journal Software Engineering and Computer Science (IJSECS), 4*(3), 1261–1279. https://doi.org/10.35870/ijsecs.v4i3.3073

Sudrajat, J., Maulana, A. I., Riswaya, A. R., Marjito, M., & Rizqiati, F. (2025). Customer relationship management information system analysis using RFM dashboard for customer segmentation strategy (Study on a ritel fashion company in Bandung). Jurnal Computech & Bisnis, 19(1), 74–82. https://doi.org/10.56447/jcb.v19i1.7

Susilowati, D., Hairani, H., Lestari, I. P., Marzuki, K., & Mardedi, L. Z. A. (2022). Segmentasi lokasi promosi penerimaan mahasiswa baru menggunakan metode RFM dan K-Means clustering. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 21*(2), 275–282. https://doi.org/10.30812/matrik.v21i2.1542

Verda, D., & Muselli, M. (2025). Alternative support threshold computation for market basket analysis. AppliedMath, 5(2), 71. https://doi.org/10.3390/appliedmath5020071

Yıldız, E., Güngör Şen, C., & Işık, E. E. (2023). A hyper-personalized product recommendation system focused on customer segmentation: An application in the fashion retail industry. *Journal of Theoretical and Applied Electronic Commerce Research, 18*(1), 571–596. https://doi.org/10.3390/jtaer18010029

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.