

Descriptive data mining for multi-shelf product allocation in traditional retail

Singgih Saptadi, Wiwik Budiawan, Ary Arvianto, Purnawan Adi Wicaksono,
Chaterine Alvina Prima Hapsari, Dhimas Wachid Nur Saputra

Department of Industrial Engineering, Faculty of Engineering, Diponegoro University, Semarang, Indonesia

Article Info

Article history:

Received Dec 10, 2024

Revised May 29, 2025

Accepted May 23, 2026

Keywords:

Customer preferences
Descriptive data mining
Shelf space allocation
Shopping pattern
Traditional retail

ABSTRACT

The business expansion of minimarkets in small cities is one of the serious threats to the sustainability of traditional retail businesses or small independent retailers. Many traditional retailers eventually closed due to their inability to maintain competitiveness, as customers increasingly prefer shopping at modern retail outlets. A well-organized store layout can improve the shopping experience of customers, which has an impact on customer satisfaction and retail competitive advantage. Currently, shelf space allocation in traditional retail is still inattentive, making the placement of products on the shelf random and erratic. Based on these problems, this research aimed to design multi-shelf product allocation according to customer shopping patterns by combining clustering algorithms and market basket analysis (MBA). Clustering aims to divide data points into two different clusters, namely dominant product and less favored product, while MBA aims to identify the customer purchase pattern and preferences. The three MBA scenarios produced four, twelve, and forty rules. The research successfully designed two layouts by utilizing a combination of clustering and MBA algorithms. The utilization of data mining allows traditional retailers to extract information from the database to be arranged into a layout design that fits the shopping patterns and customer preferences.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Dhimas Wachid Nur Saputra

Department of Industrial Engineering, Faculty of Engineering, Diponegoro University

Prof. Jacub Rais Street, Tembalang, Semarang, Indonesia

Email: dhimas.nur44@gmail.com

1. INTRODUCTION

Business expansion by modern retail in small cities is one of the serious threats to the sustainability of traditional retail businesses or small independent retail [1]. Traditional retailers initially dominated the market in small towns, but finally had to compete with minimarkets. Minimarkets that are classified as modern retail have a better business system than traditional retail, making it difficult for traditional retail to overcome business competition [2], [3]. Many traditional retailers eventually closed because they were unable to compete with minimarkets. For example, in Indonesia, based on data from the United States Department of Agriculture (USDA), as many as 39,750 traditional store outlets were no longer operating in 2022. In contrast, minimarket outlets are predicted to grow by 1,000 outlets/year [4]. Another impact that arises is a decrease in turnover and the number of traditional retail customers [5].

The failure of traditional retail in maintaining business is influenced by changes in customer shopping preferences that prefer shopping in modern retail. Research from Irfan *et al.* [2] and Ye *et al.* [6] states that retail store layout plays an important and fundamental role as a factor influencing customer shopping preferences. A well-organized store layout can improve the shopping experience of customers,

which has an impact on customer satisfaction and retail competitive advantage [7], [8]. In addition, a well-planned layout can provide comfort for customers, so customers spend longer browsing the store, increase the likelihood of impulsive buying, product sales, and encourage customers to shop again [2], [8], [9].

Modern retailers design the layout based on category areas such as clothing and food, which makes shelf space allocation is not a major concern. Meanwhile, traditional retailers with more limited space require detailed layout, including shelf product allocation strategies. Shelf space allocation is often based on short-term operational needs rather than strategic planning, making it occasionally less relevant. For example, when instant food products are sold out, empty shelves are refilled with products from other categories. In some cases, instant food is placed near body care products, making the placement less relevant due to the category mismatch. These conditions decrease customer satisfaction due to changes in layout that make it difficult to find products.

As a result, customers tend to switch to modern retailers that offer more organized layouts. These arbitrary and unstructured placements also lead to a significant gap in product sales. Product displays are not maximized as a tool to introduce the available products and become ineffective as a medium for product awareness. The significant differences in product variety and assortment between traditional and modern retail formats prevent traditional retailers from adopting modern retail layouts [10]. Therefore, traditional retailers face the challenge of developing customized layout and shelf arrangement strategies that not only address the limitations of their space but also enhance product visibility and customer satisfaction, both of which are currently hindered by arbitrary and unstructured placement.

Research related to retail layout has been extensive and utilizes various approaches such as agent-based simulation [11], data mining [12], [13], and artificial intelligence [14]. Data mining extracts information on customer purchase patterns through customer transaction datasets by modeling association rules [15]. Research by Febrianti *et al.* [16], Fageeri *et al.* [17], Ünvan [18], and Martinez *et al.* [19] employed association rule mining (ARM) modeling to extract information from customer transaction databases. Based on these works, ARM modeling can be applied at various product levels, ranging from category and subcategory to individual product levels.

However, while these studies successfully explored and generated association rules, they tend to focus more on the exploration of these rules rather than proposing concrete steps for practical implementation. On the other hand, Kuşakcı and Cesur [20] and Yanti *et al.* [13] developed an ARM model that results in a macro-level retail store layout design with concrete steps. However, this design does not address the micro-level layout or shelf space allocation. The existing research has not yet provided a comprehensive model that integrates macro-level layout planning with detailed shelf space allocation (micro-level layout), which remains a challenge for traditional retailers.

This study aims to develop a multi-shelf space allocation model based on customer shopping patterns to address layout arrangement issues in traditional retail at both macro and micro levels. The research utilizes two algorithms, namely clustering and market basket analysis (MBA) (descriptive data mining). Clustering is used to identify product groups that dominate sales and are less favored, considering the uneven sales trends conducted in traditional retail. The uneven product sales resulted from suboptimal product placement, which failed to maximize product visibility and attract customer interest. The results of this clustering model will affect the rules formed through MBA modeling. If the rules formed contain two products that are included in a group of products that dominate sales, the rules will be eliminated, and the products will not be placed close together. There are three scenarios used to create models with the output of multi-shelf product allocation design. After the rules are formed, layout and shelf allocation are planned using activity relationship chart (ARC). The research results can be a solution to help traditional retailers solve shelf product allocation problems.

The rest of the paper is organized as follows: section 2 provides a literature review on the application of data mining for retail layout design. Section 3 details the research methodology used in this study. Section 4 presents the research findings, including the proposed layout design and its analysis; and finally, section 5 concludes the study and outlines potential directions for future research.

2. DATA MINING FOR RETAIL LAYOUT

Retail can be categorized into two types: traditional retail (small-independent retail) and modern retail. Traditional retail refers to a home-based business that is usually run by one family and with small selling space. The characteristics of traditional retail include being independent, having fewer than five employees, targeting customers in the surrounding community, and tend to be end-users, relying on independent funds for capital, serving customers' daily needs, and usually being available only in offline stores. Traditional retail also typically experiences unplanned cash flow and lacks information systems to support store performance.

On the other hand, modern retail refers to corporate-chain businesses that operate multiple branch stores and typically employ more than five people. They target a wider customer base, from end-users to small-scale retailers, and fulfill customers' needs ranging from daily to monthly items. These retailers usually have both online and offline stores and are equipped with well-planned information systems and cash-flow management [10].

The competition in the retail sector forces businesses to constantly develop innovations and strategies to increase competitive advantage [7]. One of the widely used tools is the extraction of dataset information through data mining modeling [21]. In the retail sector, data mining can be used to predict customer shopping habits, MBA, product placement, customer segmentation, and others [16], [21], [22]. In terms of layout planning, association rules algorithm modeling can be used to discover hidden relationships between categories or products [13]. Clustering algorithm modeling is widely used to develop marketing strategies by generating customer segments [22] (e.g., [23]–[25]) and product segmentation (e.g., [26]–[28]). Meanwhile, MBA can be used to develop marketing strategies and improve store layouts [13], [18]–[20]. Research by Parikh and Abdelfattah [24] compared four types of clustering models combined with the recency, frequency, and monetary value (RFM) model to group customers. The model formed two clusters that can be used as a guide for developing a differentiation strategy for supermarkets.

Research by Golderzahi and Pao [25] combines clustering and regression algorithms to predict revenue from retail businesses. The clustering algorithm is used to group customers, and the modeling results become additional attributes in regression modeling. The results showed that the combination of the two algorithms had better performance in predicting revenue. In addition to grouping customer segments, clustering can also be used to group products. Research from Gustriansyah *et al.* [26] and Ruitenbeek *et al.* [27] forecasted sales by combining clustering and forecasting methods. Research by Tabianan *et al.* [28] attempts to group products into two clusters representing the most popular and desirable products.

Research on data mining in layout planning primarily focuses on the placement of shelves or departments of each product category and excludes the allocation of products on each shelf [13]. The reason is that most studies have used data from modern retailers with large selling spaces, such as hypermarkets and supermarkets. For example, Yanti *et al.* [13] conducted a study to develop retail sales strategies using the ARM model. The results of this modeling can be used to improve store layouts, discount strategies, and bundled offers.

Liao and Tasi [29] utilized clustering algorithms, classification and regression trees (CART), and association rules to propose alternative store layouts, bundling sales, and management development. The output of modeling can serve as a useful reference in finding potential store layouts, designing possible brand/products cross-selling bundling, and proposing effective promotional activities to get higher profits from the mining process. Kuşakcı and Cesur [20] conducted research to find a new approach to designing facility layouts through combining ARM and modified computerized relative allocation of facilities technique (MCRAFT). However, these studies are limited to modern retail, while traditional retail, characterized by smaller selling spaces and significant challenges in sustaining operations, is rarely considered the focus of research.

This research contributes by taking traditional retail as the main object and integrating two data mining algorithms to generate more structured layouts, and the multi-shelf product allocation plan includes shelf layout in the store (macro-level layout) to product placement on each shelf (micro-level shelf allocation). Furthermore, the use of the ARC diagram for layout design enables a more systematic modeling of product relationships compared to manual arrangement based on association rules. This approach provides a more practical and relevant solution for retailers with limited selling space.

3. METHOD

The research applied the cross-industry standard process for data mining (CRISP-DM) framework. CRISP-DM, which consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [30]. The business understanding stage identifies business problems and objectives. The data understanding stage is focused on identifying data patterns and determining the type of modeling relevant to business understanding. It includes collecting, describing, exploratory analysis, and checking data quality.

Next, the data preparation process involves cleaning and formatting the data to ensure clean data and data formats that are suitable for modeling. The modeling stage includes model selection and the construction of models to problem the solving. After the model is built, an evaluation or review of the modeling results is carried out to determine whether the model has addressed the business problems. The final stage is deployment, which involves writing the final report [31]. CRISP-DM is the most frequently used method as it

provides steps that can maximize research success and minimize errors [32]. Figure 1 shows the stages of research with the CRISP-DM framework.

3.1. Business understanding and data understanding

The purpose of this research is to design a multi-shelf product allocation model based on descriptive data mining to increase competitive advantage, enabling traditional retail to survive in business competition. The study utilized customer transaction datasets from traditional retail, collected from January to February. A total of 6,964 transaction records were collected. The attribute descriptions of the dataset can be seen in Supplementary Table S1. The visualization of the dataset analysis can be seen in Supplementary Figure S1. Based on the exploration results, it is observed that traditional retail has a high gap in each attribute. The sales gap of these product categories explains why traditional retailers experience losses, causing low-demand products to be stored in warehouses and eventually expire before being sold. Therefore, clustering modeling is required to identify product groups from the dataset.

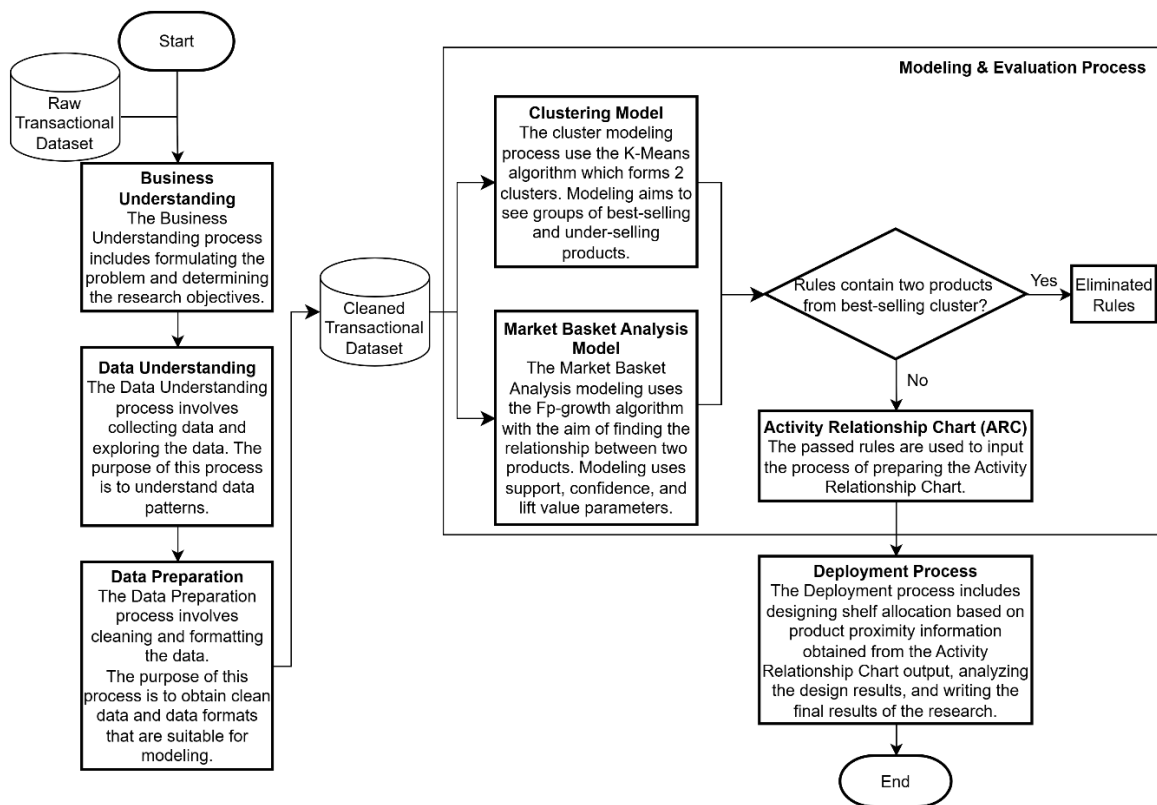


Figure 1. Research framework

3.2. Data preparation

The data preparation consists of data cleansing and data transformation. In the data cleansing process, it is known that the data has no anomalies and missing values. Then the data used can proceed to the data transformation process. The limitations of the existing dataset cause traditional retailers unable to compile association rules modeling based on the product level.

However, these limitations can be solved by modeling MBA at the product category level [16]. The dataset is transformed into a product category level that matches the product hierarchy in Supplementary Figure S2 [33]. Then, the input data format for clustering and MBA modeling can be seen in Supplementary Table S2 and Supplementary Table S3. In the data preparation process for clustering modeling, the data normalization process is carried out with the min-max normalization method, as shown in (1). Let x' be the normalized value, x be original value, x_{max} and x_{min} be the maximum and minimum value.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.3. Modeling

This research utilizes clustering algorithm modeling to cluster products and association rules to discover the relationship between two products. Clustering modeling uses agglomerative and k-means clustering algorithms and uses RapidMiner as the modeling tool. Based on preliminary findings in data exploration, clustering modeling tends to produce the optimum number of clusters (k) at two clusters. Therefore, to validate the preliminary findings, an agglomerative clustering modeling process was conducted. The modeling and results are shown in Supplementary Figure S3. Based on the results of agglomerative clustering modeling, the number of clusters (k) selected is between one and two. To avoid losing details of the data structure, the optimal number of clusters is two. In addition, by using the value of $k=2$, the initial purpose of modeling is answered. Two product groups are formed with significant group differences. Agglomerative clustering modeling with RapidMiner cannot identify the characteristics of the formed clusters, so a k-means clustering model is required. K-means clustering aims to divide data points into k distinct clusters, where each point is assigned to the cluster whose mean is closest to it. Formulas (2) and (3) show the calculation of k-means clustering. Figure 2 shows the RapidMiner-based modeling design. Figure 2(a) shows the final clustering model. Measure type and divergence in k-means parameters are set the same as the agglomerative clustering model, so both models produce the same output.

$$G(c_k) = \sum_{\vec{x}_i \in c_k} \|\vec{x}_i - \mu_k\|^2 \quad (2)$$

$$G(E) = \sum_{k=1}^K \sum_{\vec{x}_i \in e_k} \|\vec{x}_i - \mu_k\|^2 \quad (3)$$

Where, c_k is cluster number and μ_k is centroid of cluster c_k .

The MBA modeling parameter values of support (α), confidence (β), and lift value are shown in (4) to (6). The support value (α) is the number of transactions containing data items A and B from the total transactions. The confidence value (β) is the transaction proportion that contains A and B [34]. Meanwhile, a lift ratio value ≥ 1 means that the rules generated by the model are valid or their occurrence is more frequent than expected. The greater the lift ratio value, the more reliable the rules are [35].

$$\alpha(A \Rightarrow B) = \frac{|A \cup B|}{\text{total transactions}} \quad (4)$$

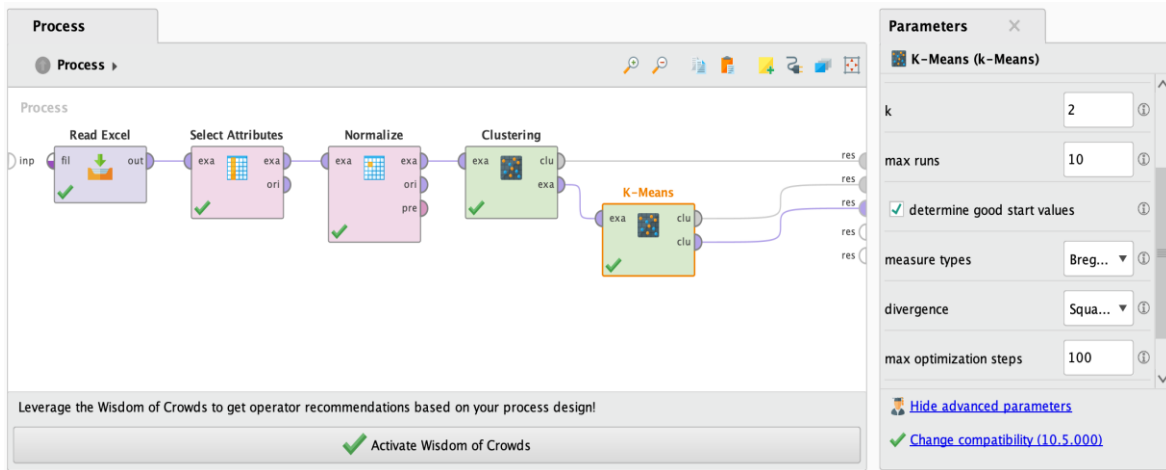
$$\beta(A \Rightarrow B) = \frac{|A \cup B|}{|\text{total transactions containing A}|} \quad (5)$$

$$\text{lift value} = \frac{\beta(A \Rightarrow B)}{\alpha(B)} \quad (6)$$

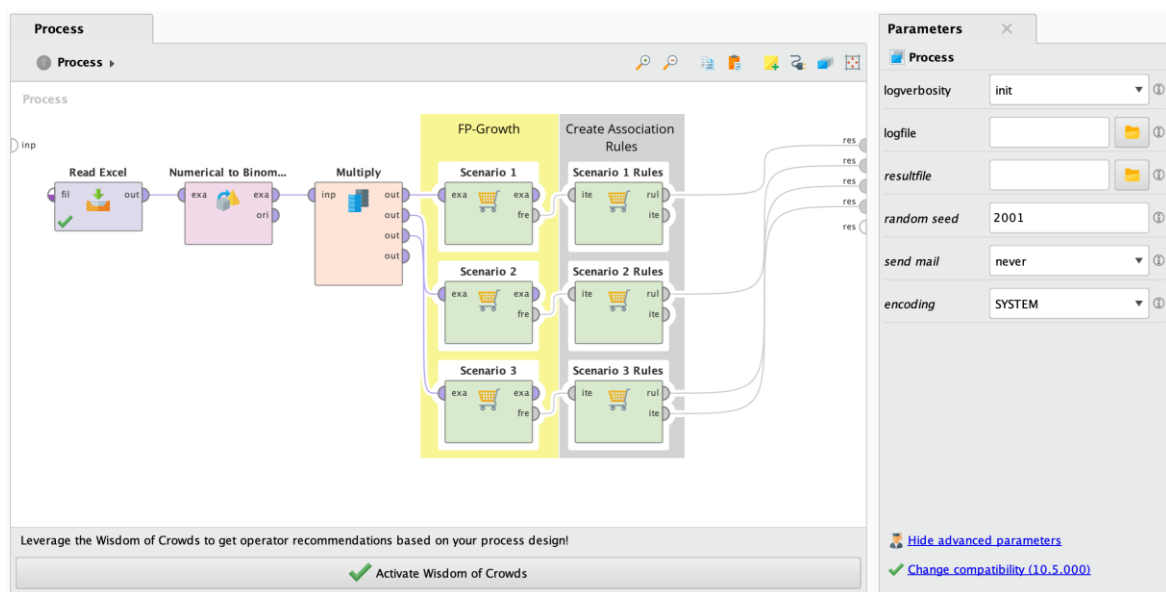
Three scenarios are used in MBA modeling: the first scenario ($\alpha = 3\%$ and $\beta = 30\%$), the second scenario ($\alpha = 2\%$ and $\beta = 20\%$), and the third scenario ($\alpha = 0.34\%$ and $\beta = 20\%$). The minimum value of support and confidence for modeling is quite low because traditional retail is dominated by single-product purchases. This dominance is due to the function of traditional retail as a fulfillment of customers' daily needs. Figure 2(b) shows the MBA model using RapidMiner. The minimum value α is set in the parameters in the frequent pattern growth (FP-Growth) operator (scenario 1, scenario 2, and scenario 3) while the minimum value β is set in the parameters in the create association rules operator (scenario 1 rules, scenario 2 rules, and scenario 3 rules).

3.4. Evaluation and deployment

The evaluation and deployment process includes designing shelf allocation based on product proximity information obtained from the ARC output, analyzing the design results, and writing the final results of the research [31]. The MBA modeling output is then used as the base for planning traditional retail layouts. Layout design using ARC. ARC is a method of arranging layouts by considering the level of proximity [36]. The ARC diagram has six levels of closeness and five reasons for closeness shown in Supplementary Table S4 and Supplementary Table S5 [37]. In the proximity reason section, modifications are made according to the results of modeling association rules. Reason code 1 is used if the lift value of the generated rules is >1 , and because not all rules produce a lift value >1 , reason code 2 can be used if the rules cannot fulfill reason code 1. These two reason codes can be used in conjunction with proximity levels A, E, and I. The product categories contained in the modeled rules must be close to each other.



(a)



(b)

Figure 2. RapidMiner-based modeling design for (a) clustering and (b) MBA

4. RESULTS AND DISCUSSION

This research focuses on planning layout and shelf space allocation in traditional retail using descriptive data mining. Descriptive data mining algorithms consist of clustering and MBA. Several previous studies have explored similar topics involving conducted similar studies involving ARM algorithms. These studies aim to explore the rules that are formed and then become the base for improving product layout [13], [18]–[20]. The research was not concerned with the possibility of forming groups of products that dominate sales. Based on the exploration of the dataset collected from traditional retailers, there is a considerable gap in terms of income, frequency of product purchases, and number of products sold between product categories. The smallest income of a product category in the dataset is Rp 91,000, while the largest income of a product category in the dataset is Rp 41,222,500. This clustering model makes product placement more optimized. Placing the products that form the rules resulting from MBA modeling close to each other can increase the probability of products being sold [15], [38]. If dominant products are placed close to each other, the visibility of unpopular products decreases, which has an impact on decreasing the probability of products being sold. Therefore, this research adds a clustering model to identify product groups and optimize the rules generated by the MBA model as a basis for preparing store and shelf layouts. There are two clustering models used in this study. Based on agglomerative clustering modeling, it is known that the optimal number of clusters to be identified is two clusters. Next, the two clusters are identified using the k-means clustering model. Figure 3 shows a plot of the k-means modeling results, where Cluster_1 is a product with a greater

number of product sales, product purchase frequency, and income compared to Cluster_0. Based on the plot, it is known that each attribute significantly differentiates between clusters.

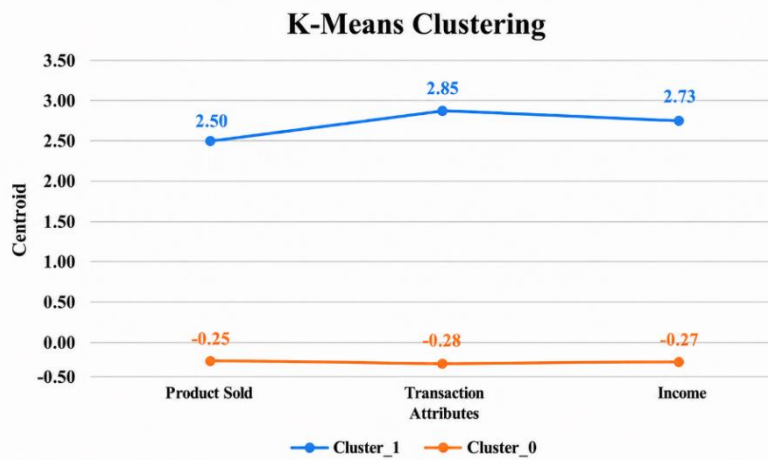


Figure 3. Clustering output

Based on the descriptive statistics presented in Table 1, it is evident that each attribute of the Cluster_1 group has a mean value smaller than the median value, indicating that most product sales, transaction frequency, and income in this group are concentrated at higher amounts. In contrast, each Cluster_0 attribute has a mean value that is greater than the median value, suggesting that most product sales, transaction frequency, and income in this group are lower. Furthermore, when examining the minimum and maximum values, the minimum value of each attribute in Cluster_1 is higher than the maximum value of the corresponding attribute in Cluster_0. Additionally, a plot of each attribute is provided in Supplementary Figure S4, showing a more detailed visualization of the data distribution across clusters. Based on the characteristics and pattern observed, it can be concluded that Cluster_1 is a product group that dominates sales with higher product sales, transaction frequency, and income, whereas Cluster_0 consists of products that are less favored, resulting in smaller sales, transaction frequency, and income.

Table 1. Descriptive statistic of clustering output

Statistics	Product sold (pcs)		Transaction frequency		Income (rupiah)	
	Cluster_0	Cluster_1	Cluster_0	Cluster_1	Cluster_0	Cluster_1
Minimum	84	2,631	31	1,654	91,000	10,343,500
Maximum	2,469	5,086	863	2,097	6,772,000	41,222,500
Mean	426	5,076	178	1,875	1,656,250	40,522,000
Median	707	4,264	266	1,875	2,527,467	30,696,000

Each attribute in the clustering model exhibits a positive correlation, as shown in Table 2. These correlations range from moderate to strong. Strong positive correlation indicates that an increase in one attribute leads to a proportional increase in others [39]. Additionally, a moderate positive correlation is observed within Cluster_1, particularly between the number of products sold and income. The correlation value of 0.600 suggests that while an increase in one attribute tends to be associated with an increase in others, the effect is less pronounced compared to a strong positive correlation. This variation in correlation is likely due to the significant differences in unit prices among products.

Table 2. Correlation output

Attribute 1	Attribute 2	Cluster_0	Correlation	Cluster_1	Correlation
Product sold	Transaction	0.791	Strong	0.871	Strong
Product sold	Income	0.914	Strong	0.600	Moderate
Transaction	Income	0.971	Strong	0.728	Strong

The relationship between the number of products sold and transaction frequency may indicate that product visibility influences sales [40]. Products that frequently appear in transactions have higher visibility, which may increase their likelihood of being sold. The relationship between the two attributes in Cluster_1 and Cluster_0 exhibits a strong positive correlation, suggesting that products in Cluster_1 (dominant products) should not be placed adjacent to each other. Placing dominant products adjacent to each other may enhance their visibility but could also reduce the visibility of less favored products. A decrease in visibility may lower the probability of Cluster_0 products being sold.

Based on the results of the clustering and correlation analysis, any rules generated by the MBA model that include two products from the dominant product group (Cluster_1) will be excluded. This exclusion aims to prevent a further increase in the sales of Cluster_1 products. Instead, products in Cluster_1 are expected to serve as a gateway product, attracting customers to purchase items from Cluster_0 that are placed adjacent. The MBA was modeled with three scenarios: the first scenario ($\alpha = 3\%$ and $\beta = 30\%$), the second scenario ($\alpha = 2\%$ and $\beta = 20\%$), and the third scenario ($\alpha = 0.34\%$ and $\beta = 20\%$). The modeling results are presented in Supplementary Table S6, Supplementary Table S7, and Supplementary Table S8. The first scenario produced four rules, the second scenario produced twelve rules, and the third scenario produced forty rules.

Effective product placement serves to provide comfort and satisfaction to customers [2], [8], [9]. However, a lack of knowledge and reference material often prevents traditional retailers from optimizing product layout arrangements. Unlike modern retail stores, traditional retail spaces are smaller, making product placement more challenging [10], [41]. This limitation complicates layout adjustments and shelf space allocation for products, as the number of products must be arranged within a confined area [42]. Currently, traditional retailers tend to arrange product placement randomly and erratically. Figures 4 and 5 show the layout of one of the existing traditional retailers. For example, the electronics shelf (Shelf O) is adjacent to the rice product category (Shelf P). Additionally, based on field observations in Figure 5, show that when rice and eggs are sold out, the shelves for both categories are temporarily filled with products from the powder drink and syrup category. Once rice and eggs are restocked, they are placed in any available empty space. This happens continuously, making it difficult for customers and employees to find products. Therefore, improvements in product placement are necessary to create a more organized and comfortable layout. These improvements involve analyzing customer purchasing patterns through the data mining process and transforming it into a shelf space allocation layout design.

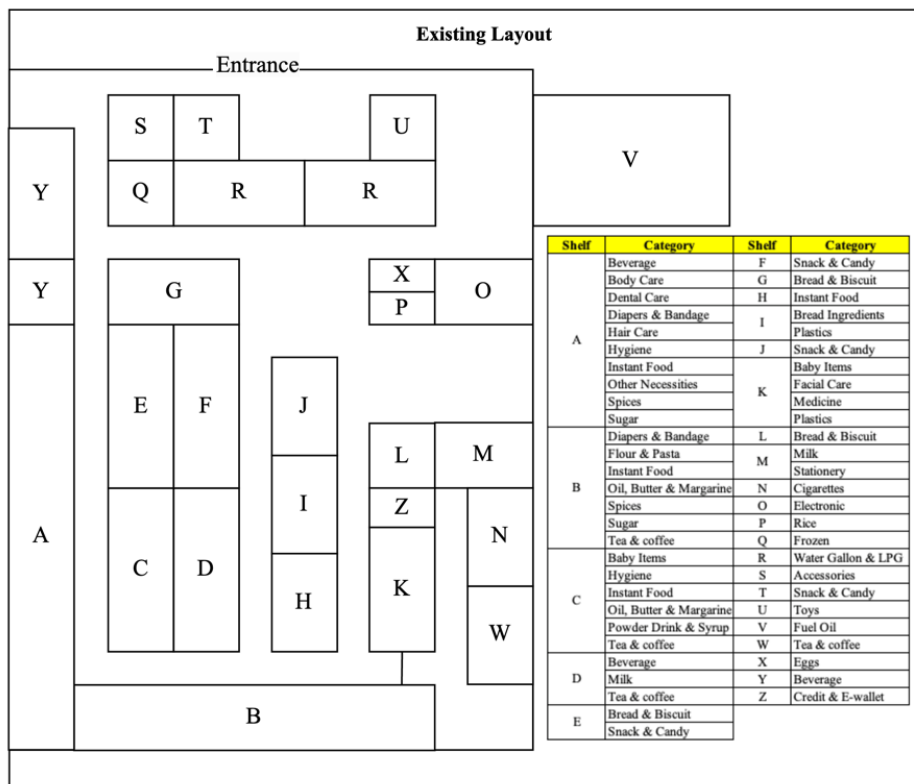


Figure 4. Existing layout



Figure 5. Existing condition of traditional retail

The first scenario generated four rules with the highest α , β , and lift ratio values in the bread and biscuit→snack and candy (5.8%; 46.5%; 1,424). This rule appears in 5.8% of the total transactions, and 46% of the transactions containing bread and biscuit category items also bought snacks and candy simultaneously in one transaction. The complete modeling results are shown in Supplementary Table S6. Based on the existing shelf layout, the four rules generated are located close to each other, making the design of shelf allocation cannot be conducted. The second scenario ($\alpha = 2\%$ and $\beta = 20\%$) generated twelve rules, with four rules also appearing in the first scenario. The detailed results of the second modeling are shown in Supplementary Table S7. In this second scenario, the rule with the highest α value remains bread and biscuit→snack and candy rule at 5.8%. Meanwhile, the highest β and lift ratio values in this scenario are observed in the sugar→tea and coffee rules of 56.2% and 6.108. The second scenario demonstrates that the small values of α and β in the modeling are not inevitably negative [43]. The primary goal of this modeling is to increase the sales of less-favored products. Thus, smaller α and β values can generate more association rules for further analysis. Overall, bread and biscuit→snack and candy are the most frequently occurring rules in the dataset. However, only 46% of the transactions containing bread and biscuit category items also included snacks and candy in the same transaction. Conversely, tea and coffee→sugar appears only 2% of the dataset, yet 56% of the total transactions containing tea and coffee items also included Sugar in the same transaction. The resulting lift ratio value is greater than 1, confirming that this rule is valid. The ARC for the second scenario layout can be seen in Supplementary Figure S5. The results of the store layout plan and space allocation product are shown in Figure 6, while detailed product allocation on each shelf can be found in Supplementary Figure S6.

The third scenario ($\alpha = 0.34\%$ and $\beta = 20\%$) was developed using the smallest α value to allow all category items to pass the first support elimination. The third scenario generated forty rules. Based on the third model, the rules with the highest α and β values are the same as those in the second scenario. The difference lies in the highest lift ratio value, which is found in the dental care→hygiene rule at 7.974. However, this rule has an α value of only 0.4% and β value of 43%. The complete modeling results are shown in Supplementary Table S8. The ARC for the third scenario layout is presented in Supplementary Figure S7. The results of the store layout plan and shelf space allocation are shown in Figure 7, while detailed shelf space allocation can be found in Supplementary Figure S8. The last scenario is the scenario with the largest layout modifications and the largest number of rules generated from MBA model. Changes affect more than 50% of the existing layout, including relocating shelves (e.g., Shelves S and U were moved adjacent to Shelf T) and reassigning shelf contents (e.g., Shelf I, which initially contained bread ingredients and plastics category products, was converted into a Bread and Biscuit category shelf). Both layout designs can be an alternative shelf space allocation product for traditional retailers.

Based on the three models, clustering and MBA in traditional retail differ from modern retail. The designed model must be adapted to the unique characteristics of traditional retail. Traditional retail primarily functions to meet customers' daily needs, leading to a dataset that is often dominated by single-item transactions. These characteristic limits the effectiveness of modeling based on product group datasets.

To address this, modeling should use category-level datasets. The output of the modeling can be used to develop product bundling strategies with the aim of increasing sales of less favored products. Additionally, the α and β value in MBA modeling should not be set too high to ensure meaningful rules.

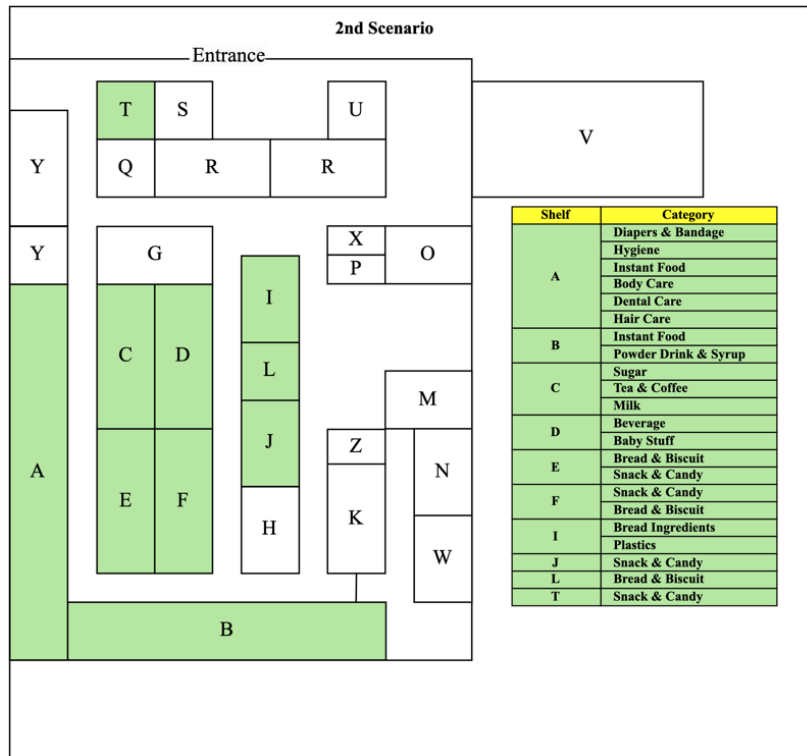


Figure 6. Layout and shelf space allocation second scenario

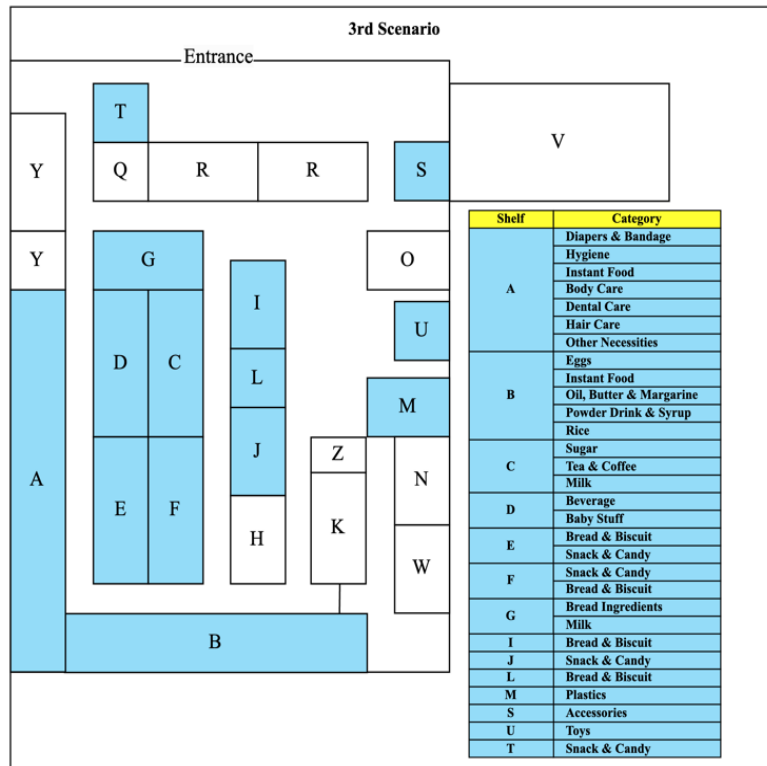


Figure 7. Layout and shelf space allocation third scenario

The model prioritizes rules with a high confidence value (β) over product occurrence (α) in transaction. A high proportion of simultaneously purchased products suggest a greater potential to attract customers. For example, the rule toys→snack and candy have a α value of 0.8%, which is considered low. However, it has a significantly higher β than other rules. Toys→snack and candy generate a β value of 52.3%, which means that 52.3% of customers who purchase toys also buy snack and candy at the same transaction. This rule was not observed in the first and second scenario modeling because the minimum α threshold was set higher than the actual α value of the rule. From the generated rules, customer purchasing patterns can be analyzed as a foundation for layout planning, shelf space allocation strategies, and strategies for introducing other products to customers to increase sales of other products that are less in demand by customers. Furthermore, all rules generated from the modeling show a positive relationship as evidenced by $\beta(X \rightarrow Y) > \alpha(X)$. This finding aligns with the study by [43], which states that when $\beta(X \rightarrow Y) > \alpha$, product X depends on product Y. This suggests that placing product X near product Y can enhance sales for both items. In this context, product Y acts as a “trigger” product, encouraging customers to purchase product X due to their strong association.

Although the developed model has successfully addressed the complexities of shelf space allocation in traditional retail, it has yet to incorporate layout optimization, price variations across categories, and the associated costs. The optimization of shelf space allocation will be the next research focus. Several studies have explored shelf allocation optimization models, such as research from Chu [44], which developed a generic algorithm to maximize the profit from shelf allocation. proposed an integrated non-linear integer problem (NLIP) model for shelf space allocation and replenishment. The model aimed to determine the optimal shelf size for each product, thereby optimizing shelf space management and increasing sales and profitability for retailers [45]. More recently, Gecili and Parikh [46] developed a shelf design and product allocation optimization model by combining particle swarm optimization (PSO) for partitioning planogram areas and constraint programming (CP) for solving the shelf sizing and product placement subproblems. Further research will enhance the current research model with an optimization model, enabling it to serve as a decision-making tool for shelf space allocation design that maximizes retail profitability.

5. CONCLUSION

This study finds that traditional retailers have the same potential as modern retailers to develop more structured and strategic product shelf allocation. By employing simple algorithms such as clustering and MBA (descriptive data mining), retailers can optimize their shelf space. Clustering identified two distinct product groups: dominant products (Cluster_1) and less favored products (Cluster_0). Correlation analysis revealed the importance of clustering in identifying product groups for a more structured layout and shelf space allocation. A strong positive correlation was found between transaction frequency and product sales, indicating that higher product visibility leads to increased sales. While placing dominant products (Cluster_1) together can boost sales, it may reduce visibility and sales of less popular products (Cluster_0). Therefore, identifying product groups is crucial for optimal layout and shelf placement. The MBA model was developed with three scenarios, with the output of each scenario being four rules, 12 rules, and 40 rules. The output was evaluated based on the clustering model results. Since dominant products (Cluster_1) cannot be near each other, rules containing two products from the dominant product group were eliminated. This research also reveals that a small support value limit does not always produce less useful rules. In the third scenario, more rules with a high confidence value and a lift value (>1) were generated, indicating that product pairs with low sales have the potential to grow when the consequent product (Y) is placed near the antecedent product (X). These findings have important implications for the retail industry, particularly in layout and shelf space allocation to enhance customer experience, customer satisfaction, and loyalty. Additionally, this study provides practical contributions to the field of retail analytics by demonstrating the effectiveness of simple yet powerful data mining techniques in improving store performance. However, this study has certain limitations, and further studies are needed to analyze the performance of the layouts and examine the actual impact on customer behavior and sales performance. In addition, future research can utilize more granular data from various retail types to provide a more in-depth analysis of customer shopping behavior.

FUNDING INFORMATION

The authors would like to thank the Faculty of Engineering, Diponegoro University, for funding support for the publication. The authors received financial support for the research, authorship, and/or publication of this article from the Faculty of Engineering, Diponegoro University. Grant Number: 26/UN7.F3/HK/IV/2023.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Singgih Saptadi	✓	✓		✓	✓				✓			✓		✓
Wiwik Budiawan	✓			✓	✓	✓			✓			✓		
Ary Arvianto		✓					✓	✓		✓				
Purnawan Adi		✓				✓	✓		✓					✓
Wicaksono														
Chaterine Alvina		✓				✓	✓			✓				
Prima Hapsari														
Dhimas Wachid Nur Saputra	✓	✓	✓	✓	✓			✓	✓	✓			✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY




The data that support the findings of this study are available from the corresponding author [DWNS], upon reasonable request. Supplementary files can be accessed through the following link: <https://github.com/dhimaswachid/Supplementary-File>.

REFERENCES




- [1] B. Kin, "Less fragmentation and more sustainability: how to supply nanostores in urban areas more efficiently?," *Transportation Research Procedia*, vol. 46, pp. 117–124, 2020, doi: 10.1016/j.trpro.2020.03.171.
- [2] W. Irfan, D. A. Siddiqui, and W. Ahmed, "Creating and retaining customers: perspective from Pakistani small and medium retail stores," *International Journal of Retail & Distribution Management*, vol. 47, no. 4, pp. 350–367, Apr. 2019, doi: 10.1108/IJRDM-03-2018-0045.
- [3] A. Khan, S. Ahmed, and F. Arshad, "Changing preference aspects from traditional stores to modern stores," in *Proceedings of the Thirteenth International Conference on Management Science and Engineering Management (ICMSEM 2019)*, 2020, pp. 694–704, doi: 10.1007/978-3-030-21255-1_53.
- [4] N. Yuningsih, "Retail foods annual," apps.fas.usda.gov. Accessed: Jun. 29, 2024. [Online]. Available: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Retail%20Foods%20Annual_Jakarta_Indonesia_ID2025-0036.pdf
- [5] R. Hensel, R. Visser, A. Overdiek, and E. Sjoer, "A small independent retailer's performance: influenced by innovative strategic decision-making skills?," *Journal of Innovation & Knowledge*, vol. 6, no. 4, pp. 280–289, Oct. 2021, doi: 10.1016/j.jik.2021.10.002.
- [6] Y. Ye, Y. Yang, and Q. Huang, "Identifying and examining the role of pop-up store design: a mixed-methods study," *Journal of Retailing and Consumer Services*, vol. 75, Nov. 2023, doi: 10.1016/j.jretconser.2023.103503.
- [7] P. Chatzoglou, D. Chatzoudes, A. Savvidou, T. Fotiadis, and P. Delias, "Factors affecting repurchase intentions in retail shopping: an empirical study," *Heliyon*, vol. 8, no. 9, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10619.
- [8] S. D. Astarini and C. Utomo, "Observation scale of space value for measuring successful shopping malls, a case study in Indonesia," *Journal of Asian Architecture and Building Engineering*, vol. 24, no. 1, pp. 115–138, Jan. 2025, doi: 10.1080/13467581.2023.2287208.
- [9] F. B. Abdelaziz, B. Maddah, T. Flamand, and J. Azar, "Store-wide space planning balancing impulse and convenience," *European Journal of Operational Research*, vol. 312, no. 1, pp. 211–226, Jan. 2024, doi: 10.1016/j.ejor.2023.06.027.
- [10] B. M. Sopha, A. Arvianto, and B. Tjahjono, "Survival strategies of traditional retailers during the COVID-19 pandemic: some insights from a developing country," *Journal of Industrial Engineering and Management*, vol. 15, no. 2, Mar. 2022, doi: 10.3926/jiem.3698.
- [11] E. Pantano, G. Pizzi, E. Bilotta, and P. Pantano, "Enhancing store layout decision with agent-based simulations of consumers' density," *Expert Systems with Applications*, vol. 182, Nov. 2021, doi: 10.1016/j.eswa.2021.115231.
- [12] G. S. Edirisinghe and C. L. Munson, "Strategic rearrangement of retail shelf space allocations: using data insights to encourage impulse buying," *Expert Systems with Applications*, vol. 216, Apr. 2023, doi: 10.1016/j.eswa.2022.119442.
- [13] R. Yanti, P. N. Maradjabessy, Q. Qurtubi, and I. P. Rachmadewi, "Determining the retail sales strategies using association rule mining," *International Journal of Advances in Applied Sciences*, vol. 13, no. 3, Sep. 2024, doi: 10.11591/ijaas.v13.i3.pp530-538.
- [14] K. Nguyen, M. Le, B. Martin, I. Cil, and C. Fookes, "When AI meets store layout design: a review," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5707–5729, Oct. 2022, doi: 10.1007/s10462-022-10142-3.

- [15] I. Cil, "Consumption universes based supermarket layout through association rule mining and multidimensional scaling," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8611–8625, Aug. 2012, doi: .
- [16] M. A. Febrianti, Qurtubi, R. Yanti, and H. Purnomo, "Multilevel association rules on customers' buying pattern based on sales transactions: a case study in retail," *International Journal of Industrial Engineering and Production Research*, vol. 35, no. 2, 2024, doi: 10.22068/ijiepr.35.2.1967.
- [17] S. O. Fageeri, M. A. Kausar, and A. Soosaimanickam, "MBA: market basket analysis using frequent pattern mining techniques," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 5s, pp. 15–21, May 2023, doi: 10.17762/ijritcc.v11i5s.6591.
- [18] Y. A. Ünvan, "Market basket analysis with association rules," *Communications in Statistics-Theory and Methods*, vol. 50, no. 7, pp. 1615–1628, Apr. 2021, doi: 10.1080/03610926.2020.1716255.
- [19] M. Martínez, B. Escobar, G.-D. M.-Elena, and D. P. P.-Roa, "Market basket analysis with association rules in the retail sector using Orange. case study: appliances sales company," *CLEI Electronic Journal*, vol. 24, no. 2, 2021, doi: 10.19153/cleiej.24.2.12.
- [20] A. O. Kuşakcı and E. K. Cesur, "Retailer layout design: a novel hybrid approach with association rules mining and MCRAFT," *International Journal of Industrial and Systems Engineering*, vol. 35, no. 4, 2020, doi: 10.1504/IJISE.2020.108545.
- [21] E. Bakhshizadeh, H. Aliasghari, R. Noorossana, and R. Ghousi, "Customer clustering based on factors of customer lifetime value with data mining technique (case study: software industry)," *International Journal of Industrial Engineering and Production Research*, vol. 33, no. 1, 2022, doi: 10.22068/ijiepr.33.1.1.
- [22] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *International Journal of Information Technology*, vol. 12, no. 4, pp. 1243–1257, Dec. 2020, doi: 10.1007/s41870-020-00427-7.
- [23] R. Belwal, S. Belwal, Z. Morgan, and L. H. Al Badi, "Profiling consumers for their shopping motivations in modern retail formats in Oman," *International Journal of Retail & Distribution Management*, vol. 53, no. 1, pp. 74–93, Jan. 2025, doi: 10.1108/IJRDM-09-2023-0581.
- [24] Y. Parikh and E. Abdelfattah, "Clustering algorithms and RFM analysis performed on retail transactions," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE, Oct. 2020, pp. 506–511, doi: 10.1109/UEMCON51285.2020.9298123.
- [25] V. Golderzahi and H.-K. K. Pao, "Revenue forecasting in smart retail based on customer clustering analysis," *Internet of Things*, vol. 27, Oct. 2024, doi: 10.1016/j.iot.2024.101286.
- [26] R. Gustriansyah, E. Ermatita, and D. P. Rini, "An approach for sales forecasting," *Expert Systems with Applications*, vol. 207, Nov. 2022, doi: 10.1016/j.eswa.2022.118043.
- [27] R. E. V. Ruitenbeek, G. M. Koole, and S. Bhulai, "A hierarchical agglomerative clustering for product sales forecasting," *Decision Analytics Journal*, vol. 8, Sep. 2023, doi: 10.1016/j.dajour.2023.100318.
- [28] K. Tabianan, S. Velu, and V. Ravi, "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
- [29] S. Liao and Y.-S. Tasi, "Big data analysis on the business process and management for the store layout and bundling sales," *Business Process Management Journal*, vol. 25, no. 7, pp. 1783–1801, Oct. 2019, doi: 10.1108/BPMJ-01-2018-0027.
- [30] N. Jiménez, S. Orellana, B. M.-Olivo, W. R.-Asanza, and I. R.-Morales, "Detection of leaf diseases in banana crops using deep learning techniques," *AI*, vol. 6, no. 3, Mar. 2025, doi: 10.3390/ai6030061.
- [31] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [32] M. Singh *et al.*, "An analytical survey of difficulty faced in an online lecture during COVID-19 pandemic using CRISP-DM," *Journal of Computer Science*, vol. 19, no. 2, pp. 242–250, Feb. 2023, doi: 10.3844/jcssp.2023.242.250.
- [33] O. Narushynska, A. Doroshenko, V. Teslyuk, V. Antoniv, and M. Arzubov, "Enhancing hierarchical classification in tree-based models using level-wise entropy adjustment," *Big Data and Cognitive Computing*, vol. 9, no. 3, 2025, doi: 10.3390/bdccc9030065.
- [34] A. Telikani and A. Shahbahrani, "Data sanitization in association rule mining: an analytical review," *Expert Systems with Applications*, vol. 96, pp. 406–426, Apr. 2018, doi: 10.1016/j.eswa.2017.10.048.
- [35] A. Ghahramani, M. Prokofieva, and M. P. D. Courten, "Health promotion campaigns using social media: association rules mining and co-occurrence network analysis of Twitter hashtags," *BMC Public Health*, vol. 25, no. 1, 2025, doi: 10.1186/s12889-024-21255-5.
- [36] A. A. Permatasari, F. N. Pramandha, M. I. Karima, and E. Santoso, "Re-layout facility to minimize defects and production cost in PT. Sendanis Jaya Makmur," *IOP Conference Series: Earth and Environmental Science*, vol. 426, no. 1, Feb. 2020, doi: 10.1088/1755-1315/426/1/012166.
- [37] N. P. Jati, A. D. I. Rahayu, S. E. Salsabila, and A. 'Azzam, "Facility layout design with corelap algorithm for educational tour," *IOP Conference Series: Materials Science and Engineering*, vol. 982, no. 1, Dec. 2020, doi: 10.1088/1757-899X/982/1/012060.
- [38] G. Aloysius and D. Binu, "An approach to products placement in supermarkets using prefixspan algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 25, no. 1, pp. 77–87, Jan. 2013, doi: 10.1016/j.jksuci.2012.07.001.
- [39] T. D. Nguyen, Q. T. P. Le, M. T. T. Doan, and H. M. Bui, "Predicting total alkalinity in saline water using machine learning: a case study with RapidMiner," *Sustainable Chemistry One World*, vol. 4, Dec. 2024, doi: 10.1016/j.scowo.2024.100032.
- [40] R. Basu, J. Paul, and K. Singh, "Visual merchandising and store atmospherics: an integrated review and future research directions," *Journal of Business Research*, vol. 151, pp. 397–408, Nov. 2022, doi: 10.1016/j.jbusres.2022.07.019.
- [41] J. Lee and K. Park, "The effects of hedonic shopping values on loyalty towards small retailers: the moderating role of trust," *Journal of Retailing and Consumer Services*, vol. 76, Jan. 2024, doi: 10.1016/j.jretconser.2023.103615.
- [42] Y. Gu and Y. Wu, "Highlighting supply-abundance increases attraction to small-assortment retailers," *Journal of Retailing*, vol. 99, no. 3, pp. 420–439, Sep. 2023, doi: 10.1016/j.jretai.2023.08.002.
- [43] X. Yan, C. Zhang, and S. Zhang, "Confidence metrics for association rule mining," *Applied Artificial Intelligence*, vol. 23, no. 8, pp. 713–737, Oct. 2009, doi: 10.1080/08839510903208062.
- [44] L. K. Chu, "An efficient approach for shelf space planning and display area allocation in convenience stores," in *2009 16th International Conference on Industrial Engineering and Engineering Management*, Oct. 2009, pp. 874–878, doi: 10.1109/ICIEEM.2009.5344195.
- [45] T. Düsterhöft, A. Hübner, and K. Schaal, "A practical approach to the shelf-space allocation and replenishment problem with heterogeneously sized shelves," *European Journal of Operational Research*, vol. 282, no. 1, pp. 252–266, Apr. 2020, doi: 10.1016/j.ejor.2019.09.012.
- [46] H. Gecili and P. J. Parikh, "Joint shelf design and shelf space allocation problem for retailers," *Omega*, vol. 111, Sep. 2022, doi: 10.1016/j.omega.2022.102634.




BIOGRAPHIES OF AUTHORS

Singgih Saptadi    holds bachelor's, master's, and doctoral degrees in industrial engineering from the Bandung Institute of Technology. Since 2001, he has been a lecturer at the Department of Industrial Engineering, Faculty of Engineering, Diponegoro University, and joined the Decision Support System Laboratory. Much of his research is in the implementation of decision-making and information technology in various industries. Currently, he is active as the Head of Department of Industrial Engineering, Diponegoro University. He can be contacted at email: singgih@lecturer.undip.ac.id.






Wiwik Budiawan    graduated from Diponegoro University, majoring in Industrial Engineering. He continued his studies at the Bandung Institute of Technology. He obtained a Ph.D. from the Toyohashi University of Technology. He has received some awards and grants concerning safety transportation, accident prevention, and academic excellence during his studies. He did research on accident analysis and prevention in mass transportation. Since 2012, he has been a lecturer at the Department of Industrial Engineering, Faculty of Engineering, Diponegoro University, joined the Decision Support System Laboratory, and Japan sleep research member. He can be contacted at email: wiwikbudiawan@lecturer.undip.ac.id.






Ary Arvianto    received a bachelor's in Industrial Engineering from Diponegoro University (in 2003) and a master's degree from Bandung Institute of Technology (in 2009). Furthermore, he completed his doctoral degree from Gadjah Mada University in 2022. Since 2004, he has been a lecturer at the Department of Industrial Engineering, Faculty of Engineering, Diponegoro University, and joined the Decision Support System Laboratory. Much of his research is in the implementation of decision-making and information technology in various industries. He can be contacted at email: aryarvianto@lecturer.undip.ac.id.






Purnawan Adi Wicaksono    is a lecturer at the Department of Industrial Engineering, Faculty of Engineering, Diponegoro University who has taught since 2001. His area of expertise is operational management. He studied as an undergraduate at the University of Indonesia from 1995 to 1999. Then, he continued his master's education at the Sepuluh November Institute of Technology from 2004 to 2006. Then, he completed his doctoral education at the Sepuluh November Institute of Technology in 2018. Currently, he is active as vice dean II of the Faculty of Engineering, Diponegoro University. He can be contacted at email: purnawan@ft.undip.ac.id.



Chaterine Alvina Prima Hapsari    is a lecturer at the Department of Industrial Engineering, Faculty of Engineering, Diponegoro University who has taught since 2022. Her areas of expertise are simulation modeling, warehouse management, and logistics and supply chain management. She received his undergraduate education at Atma Jaya University, Yogyakarta, from 2010 to 2015. She continued his master's education at the National Taiwan University of Science and Technology from 2018 to 2020. Currently, she is active as a supervisor of Undip Industrial Engineering Student Affairs. She can be contacted at email: chaterinealvina@lecturer.undip.ac.id.



Dhimas Wachid Nur Saputra    is a master's student in Industrial Engineering and Management at Diponegoro University. He took his undergraduate education in Department of Industrial Engineering, Faculty of Engineering, Diponegoro University from 2018 to 2022. He has research interests in data mining and machine learning. He is currently a laboratory assistant at Decision Support System and is active in various projects with lecturers from the laboratory. He can be contacted at email: dhimas.nur44@gmail.com or dhimaswachid@students.undip.ac.id.