

Optimization of Cross-Selling Strategies and Retail Product Layouts of Building Materials Based on Market Basket Analysis Using the Apriori Algorithm

Yolanda Joice Maria Lumbantobing^{1*}, Nur Alamsyah², Titan Parama Yoga³

Sistem Informasi^{1,2,3}
Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia^{1,2,3}
<https://unibi.ac.id/>^{1,2,3}
yolanjoyce@gmail.com^{1*}, nuralamsyah@unibi.ac.id², titanparama@unibi.ac.id³

Abstract. In an era of increasingly fierce retail competition, optimizing product layout and sales strategies is a matter of urgency for businesses seeking to maintain and increase revenue. CV. Tolong Kita Jaya, a building materials store, is currently facing challenges in determining an effective product placement strategy and has not yet maximized its cross-selling potential. This is due to a lack of analysis of past consumer purchasing patterns, resulting in frequent missed opportunities to sell related products. Therefore, this study aims to address this issue by implementing a data mining approach. This study employs the comprehensive Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, ranging from business understanding to the evaluation stage. For modeling, this study implements Market Basket Analysis (MBA) using the Apriori algorithm. This model was specifically chosen to extract association rules from a set of historical transaction data to identify relationships among building materials that are frequently purchased together. The test results from this modeling revealed several product combinations with high and valid support, confidence, and lift ratio metrics. The key findings of this study provide a strategic foundation for redesigning retail layouts, in which items with strong associations are positioned in close proximity. Furthermore, these findings facilitate the formulation of actionable cross-selling strategies by management, such as the creation of promotional sales bundles, which are ultimately projected to increase the overall sales volume of building materials.

Key words: Apriori Algorithm, CRISP-DM, Cross-Selling, Dashboard, Market Basket Analysis.

1. INTRODUCTION

Public services and operations at CV. Tolong Kita Jaya in the era of digital transformation have generated a massive amount of historical transaction data. In practice, the use of this historical data is still limited to financial administrative reporting and checking outstanding customer balances. As a result, this operational data has not been utilized sustainably as an evaluation tool or to objectively identify patterns in consumer spending. The shift from merely passive administrative reporting to data-driven evaluation tools is now crucial to prevent the company from becoming "data-rich but information-poor." [1], [2]

Efforts to map out sales strategies have so far been hampered by the lack of standardized quantitative indicators, resulting in field staff providing service that remains highly reactive [3]. The absence of these guiding variables prevents the proactive identification of cross-selling opportunities, making the evaluation of marketing strategies susceptible to managerial bias [4]. Various previous studies have, in fact, implemented data mining as

a solution in the retail sector. However, the majority of these studies have focused solely on evaluating algorithm performance without integrating the findings into practical operational guidelines and spatial layouts [5], [6]. There is a methodological gap in formulating cross-selling strategies that are validated not only through conditional probabilities but also through the significance of absolute correlations between complementary items.

To address these methodological issues and gaps, this study employs a data mining approach using the CRISP-DM framework [7], [8]. This approach provides a foundation for organizational evaluation by transforming raw data into objective and measurable insights. This study integrates the processing of historical transaction data using the Apriori algorithm to extract item combinations by setting minimum support and confidence parameters. This model is then rigorously validated using the Lift Ratio metric to ensure that the recommended association rules are complementary and statistically significant [9], [10], [11].



Comprehensively, this study aims to map purchasing association rules at CV. Tolong Kita Jaya in order to formulate Standard Operating Procedures (SOPs) for cross-selling and the reorganization of product layouts. The novelty of this study lies in the integration of the Apriori algorithm's outputs into workflow-based physical layout design, informational layout strategies, and the development of an Interactive Analysis Dashboard. The contribution of this study is to transform a large volume of historical transaction data into an objective managerial decision-support tool.

2. RELATED WORK

First Various international studies have explored the application of data mining in the commercial sector to optimize operational decisions. Arora et al [12] and Febriani et al [13] applied Market Basket Analysis to predict consumer purchasing patterns, demonstrating that the Apriori algorithm is highly effective in identifying product correlations to support personalized marketing and cross-selling strategies. These findings are reinforced by the computational approaches of Pardeshi et al [14] and Hendra et al. [15], who leveraged historical shopping basket data and association rules to build accurate recommendation systems within business ecosystems.

Focusing specifically on optimizing retail layout, Ridwan and Lubis [16] utilized association rules generated by Apriori on a transaction dataset to design cross-selling strategies and determine product placement. The study confirmed that arranging products with strong correlations side by side can significantly improve sales efficiency.

Based on this literature review, the majority of international research on Market Basket Analysis still focuses on extracting mathematical evaluation metrics to generate static layout and promotional recommendations [5], [17]. There is a methodological gap regarding how to convert the outputs of these complex algorithms into interactive monitoring tools for management. Therefore, this study offers a novel approach by not only generating tactical association rules but also integrating them into an Interactive Analysis Dashboard, as well as formulating physical (workflow-based) and informational layout strategies using signage.

3. METHODS

This quantitative applied research was conducted using the Cross-Industry Standard Process for Data Mining (CRISP-DM), an industry-standard methodology. This framework was chosen to ensure that the entire data mining process does not merely result in statistical figures, but can be transformed into operational solutions in the field [18]. The research methodology adopted for this study follows the CRISP-DM framework to ensure a business-aligned data mining process, as illustrated in Figure 1.



Fig. 1. The Iterative Process of the CRISP-DM Methodology.

3.1. Problem Identification & Business Understanding

The primary focus of this research is to identify customer shopping basket patterns based on historical transaction data from CV. Tolong Kita Jaya (January 2025–December 2025). The business objectives are aligned with the goal of formulating recommendations for cross-selling strategies and adjustments to the physical store layout to optimize sales profitability [19].

3.2. Data Understanding

Strict attribute selection is enforced to protect consumer privacy and keep the algorithm focused on affinity analysis. Operational variables are reduced, leaving only two essential attributes: Order_Number, which identifies the shopping cart (Transaction ID), and Item_Name, which identifies the purchased item.

3.3. Data Preparation

To prevent algorithm performance instability caused by the quality of the raw data, the dataset is processed through a data cleansing stage. The data cleansing steps performed are presented in Table 1.

TABLE 1. Data Cleansing Processes

Attributes / Initial Conditions	Processing Methods	Transformation Notes
Missing Values (NaN)	Row Deletion	Delete transaction rows that lack a valid item name or contain administrative recording anomalies (e.g., "Mr. X's notes").
Textual Inconsistencies	Text Normalization	Apply case folding (lowercase) and trimming (removal of extra spaces) to standardize item names.
Single-Item-Basket	Filter Exclusion	Automatically exclude transactions containing a single item from the algorithm's



Attributes / Initial Conditions	Processing Methods	Transformation Notes			
		association because they commodity pairs.	calculations have no		

The final dataset used for Market Basket Analysis was derived from sales transaction records of CV. Tolong Kita Jaya. Initially, the absolute raw dataset comprised 819 transactions with 3,005 item purchasing records across 99 unique products. To ensure data quality, an initial data cleansing stage was performed by removing 1 transaction containing administrative recording anomalies (17 records of non-item notes names "Mr. X's notes), resulting in 818 valid transactions with 2,988 item records. Subsequently, to optimize association rule mining, 103 single-item transactions were excluded as they do not contribute to item co-occurrence patterns. Consequently, the cleaned final dataset consisted of 715 multi-item transactions comprising 2,885 item purchasing records used for generating association rules.

The cleaned tabular data was then transformed into a structure that could be processed by the Python programming language, with the aggregation details shown in Table 2[19].

TABLE 2. Data Format Transformation and Aggregation

Early Attributes	Processing Methods	Transformation Notes
Transaction_ID	Basket Grouping	Serves as a unique group identifier to represent a single shopping cart.
Item_Name	Binary Encoding	Converted into a binary matrix (cross-tabulation), where a value of 1 (True) indicates the presence of an item and 0 (False) indicates the absence of an item in the shopping cart.

3.4. Modelling

The modeling phase processes the binary matrix dataset using the Apriori algorithm. The first step is to extract frequent item sets using the support metric. The minimum support value is set at 0.5% (0.005) to account for correlations among minor items[20]. The minimum support threshold was set at 0.5% to capture valuable hidden patterns among less frequently purchased, yet strongly complementary, building materials. The mathematical formula is as follows:

$$\text{Support } (A \cup B) = \frac{n(A \cup B)}{N} \quad (1)$$

The second stage involves deriving association rules, which are evaluated using the conditional probability confidence metric. The minimum confidence threshold is set at 20% (0.2). A minimum confidence threshold of 20% was selected to ensure a reliable baseline probability for the rules without excessively filtering out meaningful

associations within the sparse retail dataset. The mathematical formula is as follows:

$$\text{Confident } (A \rightarrow B) = \frac{n(A \cup B)}{n(A)} \quad (2)$$

3.5. Evaluation

The strength of the association for each rule is evaluated using the Lift Ratio metric[19]. A Lift Ratio greater than 1 indicates a positive association between the items, demonstrating that their co-occurrence is higher than expected under independent conditions, though it does not imply a causal relationship, the mathematical formula is as follows:

$$\text{Lift } (A \rightarrow B) = \frac{\text{Support } (A \cup B)}{(\text{Support } (A) \times \text{Support } (B))} \quad (3)$$

Only association rules with a lift ratio greater than 1 are considered computationally valid. To improve system efficiency, rule pruning is applied using a frozenset data structure to retain the best association pattern from each unique combination, specifically by selecting the rule that yields the highest Lift Ratio.

3.6. Deployment and Interpretation

The computational knowledge validated during the evaluation phase was translated into three tactical implementations: (1) Reorganization of the physical layout of storage racks based on workflow (workflow-based layouting), (2) A strategy for placing passive visual communication media (signage) to remind staff about cross-selling heavy-tonnage materials, and (3) Development of a centralized visual interface in the form of an Interactive Analysis Dashboard as a real-time decision support system for on-site management[21].

4. RESULTS AND DISCUSSIONS

4.1. Extraction of Association Rules and Evaluation of Metrics

Based on the application of a minimum support parameter of 0.5% (0.005) and a minimum confidence parameter of 20% (0.2), the Apriori algorithm successfully extracted a set of frequent item sets. A final evaluation using the Lift Ratio parameter confirmed that all generated association rules are highly valid, with every rule recording a Lift Ratio score greater than 1. This indicates a strong causal relationship among complementary items, rather than mere statistical coincidence. The specific details of these strong causal relationships, particularly the association rules that achieved the highest Lift Ratio scores, are presented in Table 3.



4.2. Implementation of Business Strategies and Store Layout (Deployment).

Antecedent	Consequent	Support	Confidence	Lift Ratio
Steel, Wood Nails	Tie Wire	0.0073	0.75	23.5962
Steel, Cement	Tie Wire	0.0073	0.6	18.8769
Cement, Wood Nails	Tie Wire	0.0073	0.5455	17.1608
Electrical Cable	Light Switch	0.0098	0.3636	14.1645
Steel, Cement	Sand	0.0061	0.5	13.6333
elbow	water hose	0.0073	0.4286	13.4835
fitting, pipe	thinner	0.0061	0.6250	13.4539
sandpaper, oil paint	sandpaper	0.0086	0.3684	12.0547
spray paint	white	0.0073	0.2727	11.7416
squat toilet	cement			
plastering trowel	reducer	0.0073	0.2727	11.7416

Based on the table, the association rule with the highest absolute correlation significance was found in the project structural materials category, specifically in the combination rule {Wood Nails, Steel} $\rightarrow$ {Tie Wire}, which recorded a record-high absolute Lift Ratio of 23.59. In addition to rules involving large structural materials, the model also successfully identified hidden correlations (hidden patterns) across functional categories. This is evidenced by a Lift Ratio of 13.45 for the combination {sandpaper, oil paint} $\rightarrow$ {thinner}, as well as a score of 11.74 for the wet-area material combination {concrete trowel} $\rightarrow$ {reducer}. To provide a broader perspective on these findings, the overall distribution of the generated association rules based on their support, confidence, and lift ratio metrics is visually summarized in the scatter plot shown in Figure 2.

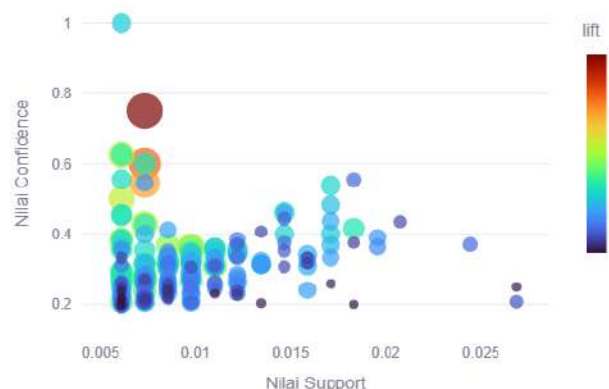


Fig. 2. Scatter Plot Visualization of Association Rule Distributions Based on Support, Confidence, and Lift Ratio Metrics

Second, the implementation pillar focused on restructuring the spatial layout of merchandise placement to address spatial inefficiencies and long product search times (lead time). In the initial layout shown in Figure 3, product arrangement at CV. Tolong Kita Jaya was still very rigid, with items grouped purely by product category. This traditional arrangement separated items that actually share a high degree of functional interdependence, forcing employees to cover inefficient distances while preparing customer orders.

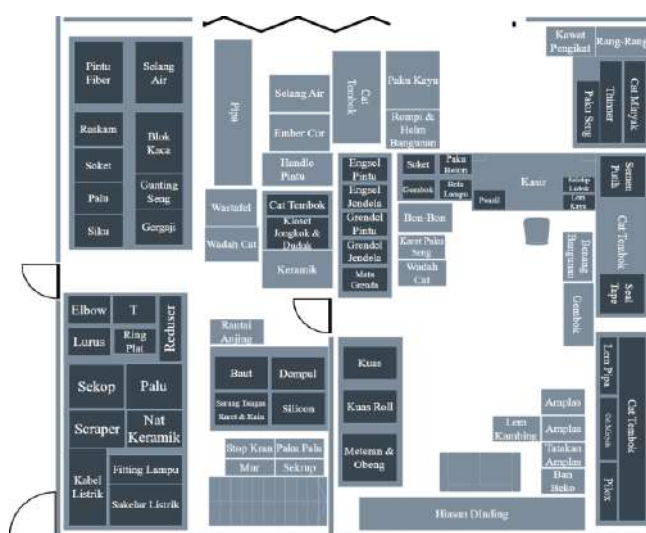


Fig. 3. Initial Store Layout of CV. Tolong Kita Jaya (Before Optimization).

The proposed layout design (Figure 4) offers a solution by implementing a workflow-based layout concept grounded in the Apriori algorithm's association clusters. Product shelves are now arranged so that items frequently purchased together such as plumbing supplies, carpentry tools, and finishing components are placed in a single functional aisle aligned with the customer's project workflow.

Third, for heavy-weight items (such as bags of cement, concrete blocks, and cast iron) that cannot be physically relocated, the company implements an informational layout strategy using visual signage. These signs serve as cross-aisle reminders for customers so they don't forget important complementary products before heading to the checkout counter.

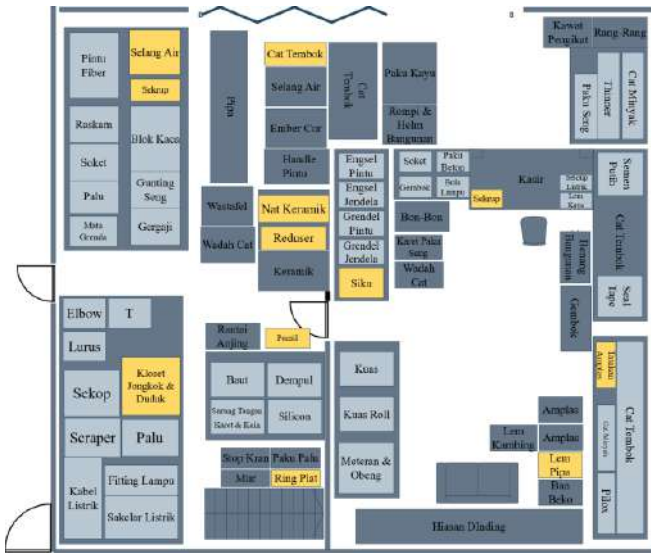


Fig. 4. Proposed Workflow-Based Store Layout Using Apriori Algorithm (After Optimization).

4.3. Development of an Interactive Analytics Dashboard

To provide a sustainable monitoring infrastructure, all computational algorithms and association rules are integrated into the web-based Interactive Analysis Dashboard software. As shown in Figure 5, this dashboard features a dedicated “Micro Analysis” module designed to allow management to explore the depth of relationships between items and translate data into operational

4.4. Discussion

While numerous prior studies have successfully applied the Apriori algorithm for Market Basket Analysis, a significant portion of this research utilizes standard data mining frameworks that primarily conclude at the rule extraction phase without integrating the findings into practical operational guidelines[5], [6]. This study addresses that methodological gap by implementing the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework. Unlike conventional approaches that stop at evaluating algorithm performance, the CRISP-DM methodology employed in this research ensures that the statistical findings are directly aligned with practical business understanding. This approach effectively bridges the gap between raw data patterns and actionable retail strategies tailored for the building materials sector.

5. CONCLUSIONS

This study successfully extracted transaction history data from CV. Tolong Kita Jaya to mitigate managerial bias by identifying objectively validated purchase association rules (Lift Ratio > 1). The implementation of the Apriori algorithm using the CRISP-DM framework did not stop at the production of statistical metrics; rather, it proved capable of serving as a strategic foundation for executing three operational pillars: proactive cross-selling mapping, workflow-based product layout reconstruction, and informational layout strategies for heavy materials. The novelty of presenting the algorithm's outputs within an

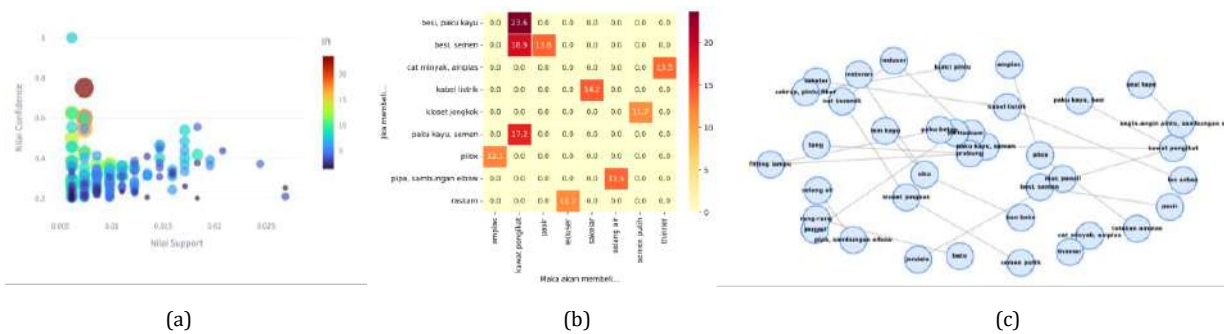


Fig. 5. Visualizations of the generated association rules: (a) a scatter plot evaluating the relationship between support, confidence, and lift metrics; (b) a heatmap matrix indicating the association strength between antecedent and consequent items; and (c) a network graph illustrating the interconnections among building material products.

strategies without requiring programming expertise. This dashboard transforms statistical metrics into three key visualizations to facilitate decision-making. First, the Scatter Plot maps rules based on Support and Confidence, where the size and color intensity represent the Lift Ratio score to instantly identify absolute correlations. Second, the Heatmap highlights the top 10 correlations as a direct guide for cashiers in executing cross-selling strategies. Third, the Spatial Network Graph displays an ecosystem of complementary products as an objective blueprint for comprehensively reorganizing the store's physical layout.

Interactive Analysis Dashboard has successfully created a ready-to-use early warning system and tactical decision-making tool. The proposed workflow-based physical layouting and informational signage are expected to potentially improve navigation efficiency and stimulate cross-selling opportunities. Furthermore, these implementations offer potential benefits for maximizing retail profitability based on the analyzed transaction patterns.

REFERENCES



- [1] A.-R. H. Tawil, M. Mohamed, X. Schmoor, K. Vlachos, and D. Haidar, "Trends and Challenges towards Effective Data-Driven Decision Making in UK Small and Medium-Sized Enterprises: Case Studies and Lessons Learnt from the Analysis of 85 Small and Medium-Sized Enterprises," *BDCC*, vol. 8, no. 7, p. 79, Jul. 2024, doi: 10.3390/bdcc8070079.
- [2] I. Pugna, D. Boldeanu, M. Gheorghe, G. Cozgarea, and A. Cozgarea, "Management Perspectives towards the Data-Driven Organization in the Energy Sector," *Energies*, vol. 15, no. 16, p. 5775, Aug. 2022, doi: 10.3390/en15165775.
- [3] M. Berkmann, M. Eisenbeiss, W. Reinartz, and N. Schauerte, "Leveraging B2B field service technicians as a 'second sales force': How service situations affect selling activity and success," *J. of the Acad. Mark. Sci.*, vol. 52, no. 3, pp. 736–761, May 2024, doi: 10.1007/s11747-023-00964-0.
- [4] S. Akter, S. Sultana, M. Mariani, S. F. Wamba, K. Spanaki, and Y. K. Dwivedi, "Advancing algorithmic bias management capabilities in AI-driven marketing analytics research," *Industrial Marketing Management*, vol. 114, pp. 243–261, Oct. 2023, doi: 10.1016/j.indmarman.2023.08.013.
- [5] A. Ivana and I. Maryati, "Implementation of Apriori Algorithm in Identifying Purchase Relationships at Bluder Cokro Pakuwon Mall," *JAIC*, vol. 9, no. 2, pp. 556–563, Apr. 2025, doi: 10.30871/jaic.v9i2.9154.
- [6] S. Saptadi, W. Budiawan, A. Arvianto, P. A. Wicaksono, C. A. P. Hapsari, and D. W. N. Saputra, "Descriptive data mining for multi-shelf product allocation in traditional retail," *Int J Adv Appl Sci*, vol. 15, no. 2, p. 830, Jun. 2026, doi: 10.11591/ijaas.v15.i2.pp830-843.
- [7] C. Wang, A. Stupina, and S. Bezhitskiy, "Online sales prediction approach using methodology of CRISP-DM," *ITM Web Conf.*, vol. 59, p. 01006, 2024, doi: 10.1051/itmconf/20245901006.
- [8] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.
- [9] C.-W. Lee and A. Ahmatang, "Market Basket Analysis Using Apriori Algorithm: Identifying Consumer Purchase Patterns for Strategic Business Decisions," *Journal of Risk & Control*, pp. 101–114, Nov. 2025, doi: 10.47260/jrc/1216.
- [10] M. Zaky Hadi, Ari Hasudungan Pratama Pasaribu, Fatin Saffanah Didin, and Lina Aulia, "A design of cross-selling products based on frequent itemset mining for coffee shop business," *tekno*, vol. 11, no. 2, pp. 286–297, Jul. 2024, doi: 10.37373/tekno.v11i2.1065.
- [11] K. A. Putri Kausara and M. T. Sulistyono, "Comparative Analysis of Association Rule Mining Algorithms Using Apriori and Eclat on Retail Transaction Data," *JAIC*, vol. 10, no. 2, pp. 1622–1630, Apr. 2026, doi: 10.30871/jaic.v10i2.12122.
- [12] Dr. Y. Arora and Dr. Y. Arora, "Market Basket Analysis using Apriori Algorithm," *IJIRCST*, pp. 62–66, May 2022, doi: 10.55524/ijircst.2022.10.3.12.
- [13] N. N. Febriani Sm, N. G. Setyoningrum, M. Lodana, D. A. A. Pertiwi, and M. A. Muslim, "Analysis and Visualization of Purchasing Pattern in Retail Product Transaction using Apriori Algorithm," *J. Inf. Syst. Explor. Res.*, vol. 4, no. 1, pp. 41–50, Feb. 2026, doi: 10.52465/joiser.v4i1.650.
- [14] A. Pardeshi *et al.*, "An Approach to Explore Consumer Behavior Patterns in Retail Markets Using Market Basket Analysis," *Appl Stoch Models Bus & Ind*, vol. 41, no. 6, p. e70057, Nov. 2025, doi: 10.1002/asmb.70057.
- [15] H. Essalmi and A. El Affar, "Dynamic Algorithm for Mining Relevant Association Rules via Meta-Patterns and Refinement-Based Measures," *Information*, vol. 16, no. 6, p. 438, May 2025, doi: 10.3390/info16060438.
- [16] R. Ridwan and H. Lubis, "Implementation of the Apriori Algorithm for Product Arrangement in a Minimarket," *GJES*, vol. 2, no. 4, pp. 184–191, Apr. 2025, doi: 10.38035/gijes.v2i4.352.
- [17] W. Widodo, E. Riswanto, Suparyanto, and H. D. Sulistianto, "Implementation of Association Rule Method for Promotion Strategy Using the Apriori Algorithm: Case Study: Snada Accessories Store," in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, Yogyakarta, Indonesia: IEEE, Jun. 2025, pp. 563–568. doi: 10.1109/ICERA66156.2025.11087287.
- [18] C. Schröder, 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.
- [19] M. Alawadh and A. Barnawi, "A Consumer Behavior Analysis Framework toward Improving Market Performance Indicators: Saudi's Retail Sector as a Case Study," *JTAER*, vol. 19, no. 1, pp. 152–171, Jan. 2024, doi: 10.3390/jtaer19010009.
- [20] A. B. Prasetyo, "Interpretable Product Recommendation through Association Rule Mining: An Apriori-Based Analysis on Retail Transaction Data," *IJIS Int. J. Informatics Inf. Syst.*, vol. 8, no. 2, pp. 67–74, Mar. 2025, doi: 10.47738/ijis.v8i2.252.
- [21] J. N. Elquthb, "Designing Planograms for Retail Shelves: A Visual Merchandising Approach Using Apriori Algorithm and K-Means Clustering of Customer Preferences," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 86–100, Jan. 2024, doi: 10.47738/jads.v6i1.375.

