
Product Bundling Strategy for Office Supplies Retailer through Association Rules Mining: Comparative Study of Apriori and ECLAT Algorithms**Ega Silfa Yuliana¹, Raymond Sunardi Oetama²**ega.silfa@student.umn.ac.id, raymond@umn.ac.id^{1,2} Department of Information System, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia

Article Information

Submitted : (date)

Reviewed: (date)

Accepted : (date)

KeywordsProduct Bundling
Strategy, Association
Rule, Apriori, ECLAT.

Abstract

Our study aims to develop an effective bundled product promotion strategy for the office supply store to boost sales. The primary challenge is comprehending which product combinations align with customer preferences and cater to their needs. We leverage the Apriori and ECLAT algorithms for consistent rule generation, revealing robust associations between product purchases. Notably, a strong positive correlation rule emerges at a confidence level of 0.8, while at 0.9, no results are found. The identical rules derived from both algorithms signify their reliability. The shop owner employs two rules for bundled products based on a minimum Lift Ratio of 1.96. The first bundle focuses on 70gsm natural paper in Folio and Quarto sizes, capitalizing on their popularity, even though customers may prefer one size. The second bundle emphasizes notebooks, often bought together but in smaller quantities than paper products, reflecting diverse customer needs and behaviors.

A. Introduction

Product bundling is a marketing strategy in which multiple individual products or services are combined and sold as a single package or bundle [1]. Instead of selling these items individually, they are presented to customers as a single, integrated bundle. The goal is to provide customers with added value, convenience, and often cost savings when they purchase the bundled package [2]. Office supplies constitute a comprehensive array of indispensable products and materials requisite for the seamless functioning of a professional workspace [3]. These supplies play a pivotal role in upholding an orderly and efficient office environment, thereby facilitating a myriad of tasks essential for documentation, communication, storage, and presentation, ultimately contributing to heightened workplace productivity and efficacy encompassing stationery items like pens, paper, and staplers, as well as office furniture and equipment such as desks and printers, along with computers and accessories, organizational tools, cleaning provisions, breakroom necessities, business and presentation materials, and safety equipment. Office supply businesses widely use The product bundling approach to encourage customers to buy a collection of related items, ultimately boosting sales and improving the overall customer experience. Product bundling allows customers to acquire a combination of offerings in one purchase, which can be more attractive and efficient than buying each item separately.

Data mining is the methodical examination of the process involved in discovering knowledge within a database [4]. One of the data mining algorithms is association rules. Association rule-mining is a common data mining method employed to examine and comprehend sizable transactional datasets to uncover distinct patterns and rules [5]. These rules unveil the associations, dependencies, and correlations between diverse items or attributes within a dataset. While notably employed in market basket analysis to identify items frequently bought together in retail settings, the utility of association rules extends to various other domains and applications.

Two popular algorithms under association rules are Apriori and ECLAT (Equivalence Class Transformation). The Apriori algorithm scrutinizes product connections by leveraging transactional data [6]. In business office supply, the approach utilized involves the Apriori algorithm, which is employed to extract customer traces [5] and assess investment opportunities [7]. Moreover, some researchers use the ECLAT algorithm as an alternative method. Some researchers use it to investigate customer shopping habits [8] or product shelf arrangement [9]. What sets this research apart from the abovementioned research on association rules for product bundling is its focus on a specific Aji's office supply store in Riau. This store's problem is identifying the appropriate product pairings for bundling and promotional purposes. Our study aims to develop a bundled product promotion strategy to boost its sales. Two association rules algorithms, Apriori and ECLAT, are employed for comparative analysis.

B. Research Methods

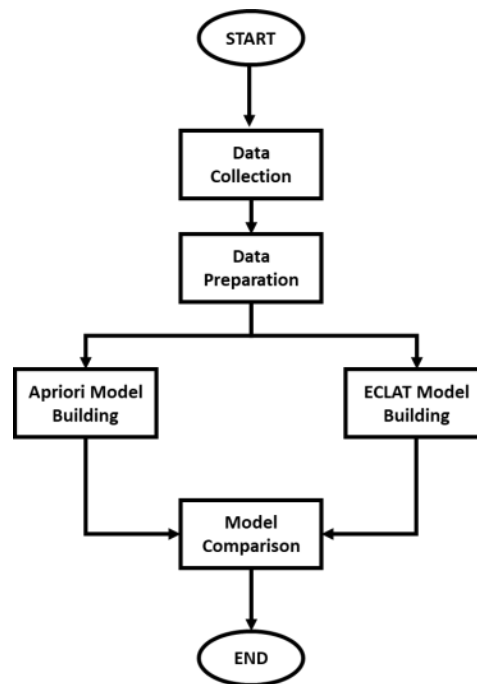


Figure1. Research Process

As depicted in Figure 1, this study commences with collecting data from the store. The data preparation process then follows it. Subsequently, two iterations of model building are executed concurrently, applying the Apriori and ECLAT algorithms. Finally, the outcomes of these two models are subsequently compared.

B.1. Data Collection

Initially, the prerequisites of the shop owner for generating association rules for product bundling are ascertained through an interview. Subsequently, sales transaction report data from the shop is compiled for two months, commencing on December 1, 2021, and concluding on January 31, 2022. Table 1 shows the data structure. It delineates a structured dataset encompassing pivotal fields: "TransactionDate" for recording purchase dates, "CustomerName" holding customer names, "CustomerAddress" storing customer addresses, "ItemCategory" indicating item categories, and "ItemNames" specifying item names.

TABLE 1. Data Structure

<i>Fields</i>	<i>Structure</i>	<i>Description</i>
<i>TransactionDate</i>	<i>date</i>	<i>Purchasing date</i>
<i>CustomerName</i>	<i>Character</i>	<i>Name of Customers</i>
<i>CustomerAddress</i>	<i>Character</i>	<i>Address of Customers</i>
<i>ItemCategory</i>	<i>Character</i>	<i>Category of Items</i>
<i>ItemNames</i>	<i>Character</i>	<i>Name of Items</i>

B.2. Data Preparation

It was essential to clean the raw sales transaction report from the store, which had irregularities in its columns and rows, to prepare the data for analysis in RStudio. The first step involved selecting the relevant attributes needed for this study: the TransactionDate, CustomerName, ItemCategory, and ItemNames. Once the data cleansing was complete, the refined dataset was saved in CSV format, ensuring compatibility with RStudio tools for further analytical procedures.

B.3. Model Building

Modeling building will be conducted with various parameter values to evaluate and compare the performance of the Apriori and ECLAT algorithms. The minimum support, confidence, minimum length, maximum length, and minimum lift ratio in each algorithm will vary. All transaction data will be used for the experiments, testing different parameter values. Minimum values of 2 will be used, and the maximum length is 3. The support will be varied from 0.1 to 1, and the minimum confidence value from 0.1 to 1. This testing aims to compare each algorithm's performance on the sales transaction data from the shop.

High-frequency pattern analysis is a technique used in data mining and machine learning to discover frequent patterns or relationships between items in a dataset [10]. It involves identifying sets of items that frequently appear together in transactions and using these sets to generate association rules. One of the critical parameters in high-frequency pattern analysis is the support value [11], which specifies the minimum frequency or occurrence threshold that an item set must meet to be considered frequent [12]. This stage finds a pattern of item combinations that meet the minimum requirements of the support value. For example, from the entire existing dataset, find the level of dominance that shows the joint occurrence at one time between item A and item B. The support value of an item is obtained by using formula (1), while the support value of the 2-item is obtained using formula (2) [13].

$$\text{Support}(A) = \frac{\text{Number of Times } A \text{ appears}}{N} = P(A), \text{Support } B = P(B) \quad (1)$$

$$\text{Support}(A \rightarrow B) = \frac{\text{Number of Times } A \text{ and } B \text{ appear together}}{N} = P(A \cap B) \quad (2)$$

Association rule Pattern Establishment is a data mining technique to discover relationships between variables in a dataset [14]. Specifically, it seeks to identify frequent patterns or sets of items that frequently co-occur in a dataset and use these patterns to generate association rules that capture the dependencies between the items [15]. After all the frequent item set patterns are found, the association rules that meet the minimum confidence requirements are searched by calculating the confidence value of the association rules $A \rightarrow B$ obtained from formula three below.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A)} = \frac{P(A \cap B)}{P(A)} = P(B|A) \quad (3)$$

The lift ratio is a measure to determine the strength of the association rules that have been formed [16]. The lift ratio value is usually used to determine whether or not an association rule has been obtained. Three possibilities will be generated when calculating the lift ratio: If the lift is worth more than 1, it is positively correlated; otherwise, if the lift is less than 1, it is negatively correlated. And if the lift result is 1, then there is no correlation, or conclusions cannot be drawn. The higher the lift value, the stronger the association or it can be said that an exciting association is associated with definite rules and has a lift value >1 [17].

$$\text{Lif}(A \rightarrow B) = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)} = \frac{P(B|A)}{P(B)} \quad (4)$$

B.4. Model Comparison

The models are compared based on the rules they generate and the time required for this process. Subsequently, the shop owner will examine the rule results for decision-making, determining which rules will be employed for product bundling.

C. Result and Discussion

C.1. Comparison of Model Results

Figure 2 presents the analysis findings on the rules generated that show the combination of the number of rules (vertical axis) and confidence (horizontal axis). The results are the same for both Apriori and ECLAT algorithms. The number of rule sets is the same for both algorithms at any combination of minimum length, support, and confidence. When minimum length was 2, and minimum support and minimum confidence were set to 0.1, a total of 25 rules were obtained, and all of them had a lift value greater than 1, indicating a positive correlation. Similar results were obtained when the confidence value was increased to 0.2 and 0.3. However, when the confidence value was lowered to 0.4, only 18 positively correlated rules were obtained.

Further decreasing the confidence value to 0.5 resulted in 17 rules, while setting it to 0.6 and 0.7 led to the discovery of only 7 and 3 rules, respectively, all of which were positively correlated. When a high confidence value of 0.8 was used, only one rule was obtained, with a lift value of 4.06, indicating a positive correlation. However, no rules were found when the confidence value was set to 0.9; the highest confidence value observed was only 82.35.

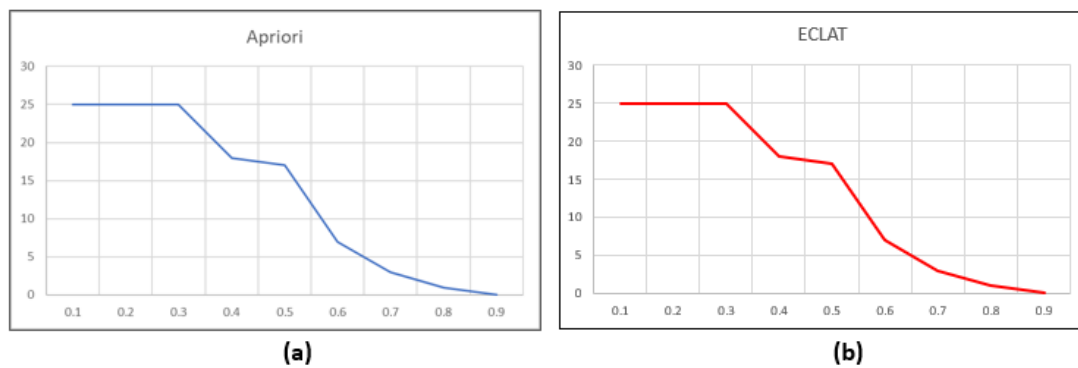


Figure 2. Comparison of Model Results between (a) Apriori and (b) ECLAT

C.2. Comparison of Rules

Figures 3 and 4 depict the outcomes of employing two distinct algorithms, Apriori and ECLAT, to uncover association rules within a dataset. In both illustrations, the vivid red highlights the rule with the most elevated lift ratio value, quantifying the correlation between the two items. Despite variations in the placement of the rules between the two figures, both algorithms have successfully pinpointed an identical set of rules. Here are ten rules generated by both algorithms:

1. When 70Gsm Folio (F4) Natural Paper is bought, likely, 70Gsm Quarto (A4) Natural Paper is also purchased.
2. Purchasing Natural 70Gsm Quarto (A4) Paper often coincides with buying Natural 70Gsm Folio (F4) paper.
3. Buying a Sidu Notebook @38 (320) and an Easy Gel Kenko Pen 0.5mm(144) frequently leads to the purchase of a Paperline Notebook @40 (280).
4. The acquisition of Paperline @40 (280) Notebook and 0.5mm(144) Easy Gel Kenko Pen is often followed by the purchase of Sidu @38 (320) Notebook.
5. Purchasing Fox Pvac 150 G(48) Glue tends to be associated with purchasing Easy Gel Kenko 0.5mm(144) Pen.
6. The Easy Gel Kenko 0.5mm(144) Pen purchase often coincides with buying Fox Pvac 150 G(48) Glue.
7. If Sidu Notebook @38 (320) and Paperline Notebook @40 (280) are bought together, likely, Easy Gel Kenko 0.5mm(144) will also be purchased.
8. When Sidu Notebook @38 (320) is purchased, it often coincides with the purchase of Paperline Notebook @40 (280).
9. Buying a Paperline Notebook @40 (280) is often followed by purchasing a Sidu Notebook @38 (320).
10. If the BPT-P Black/Blue/Red Pilot Pen is purchased, it's likely that Easy Gel Kenko 0.5mm(144) Pen will also be bought.

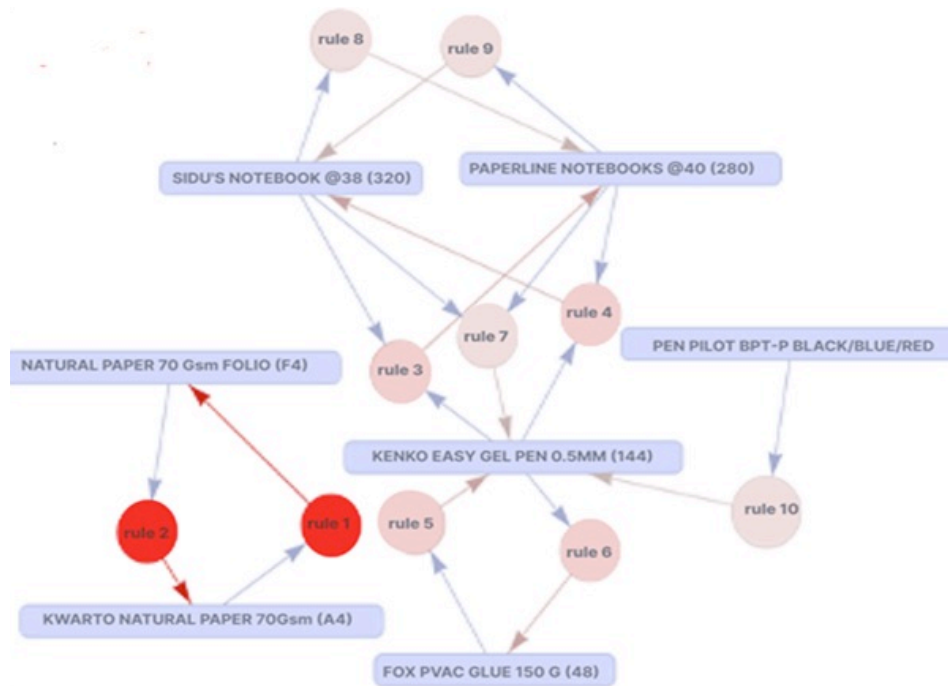


Figure 3. Apriori Rules



Figure 4. ECLAT Rules

C.3. Time Execution Comparison

Table 2 presents a comparative analysis of the execution times for model building using the Apriori and ECLAT algorithms, with the execution times being quantified in

milliseconds. The disparities in execution times between these two model-building processes are presented in the "Differences" column. The examination of this data underscores that the Apriori algorithm exhibits notably swifter execution times. As such, it can be inferred that the Apriori algorithm is the more efficient for generating association rules and is consequently recommended for research purposes. This conclusion is underpinned by the conspicuous disparity in execution times, with Apriori consistently outperforming ECLAT in speed.

TABLE 2. Execution Time Comparison between Apriori and ECLAT

<i>Confidence</i>	<i>Apriori</i>	<i>ECLAT</i>	<i>Differences</i>
0.1	46.51	52.06	-5.55
0.2	46.51	52.06	-5.55
0.3	46.51	52.06	-5.55
0.4	33.92	49.07	-15.15
0.5	38.37	55.45	-17.08
0.6	31.32	56.48	-25.16
0.7	31.32	57.33	-26.01
0.8	32.91	40.69	-7.78
0.9	30.91	38.70	-7.79

C.4. Product Bundling Decision

Among rules from both Apriori and ECLAT algorithms, the shop owner applied two rules with a minimum Lift Ratio of 1.96, as shown in Table 3. The first bundle products are made from the first and the second Rules centered around customer purchases of 70gsm natural paper in Folio and Quarto sizes. These particular products have a high level of popularity and a tendency to sell out rapidly. However, Certain customers might buy only one size at a time, driven by their immediate requirements or preferences. Even though these products are in demand, not all customers purchase both sizes simultaneously. Some may select one size over the other depending on their specific needs. The second product bundling strategy, derived from the eighth and ninth rules, focuses exclusively on acquiring notebooks. It is noted that customers often acquire notebooks simultaneously. However, the quantities purchased are not as substantial as those associated with paper products, as described in the first bundled product. It indicates that while there is a prevalent trend for customers to make concurrent notebook purchases, the volume of these acquisitions is relatively modest compared to their purchases of paper products, as exemplified in the first bundled offering.

TABLE 3. Product Bundling Decision

Rules No.	Category	Product 1	Product 2	Lift Ratio
1	Natural Paper	70Gsm Folio (F4)	70Gsm Quarto (A4)	4.06
2	Notebook	Sidu @38 (320)	Paperline @40 (280)	1.96

D. Conclusion

The Apriori and ECLAT algorithms consistently produce the same number of rules with varied confidence levels. At a high confidence level of 0.8, only one rule is generated with a strong positive correlation, while no rules are found at a confidence level of 0.9. The two algorithms revealed identical rules. These ten rules, derived from the Apriori and ECLAT algorithms, demonstrate strong associations between product purchases. While the positions of these rules may vary in the analysis figures, both algorithms consistently identify the same set of rules. In this study, the only difference between these algorithms is execution time. Apriori consistently outperforms ECLAT in speed, making it the recommended choice for association rule generation in research.

The shop owner applied two rules with a minimum Lift Ratio of 1.96 from the Apriori and ECLAT algorithms. The first product bundle includes 70gsm natural paper in Folio and Quarto sizes, popular but not always purchased together due to customer preferences. The second bundle focuses on notebooks, often bought simultaneously, although in smaller quantities than paper products. These bundles accommodate diverse customer needs and buying behaviors.

Personalize marketing campaigns based on customer behavior and cross-promote paper and notebooks to enhance marketing strategies for office stationery stores. Continuously analyze sales data to adapt marketing strategies to changing customer preferences and implement customer loyalty programs with rewards and exclusive bundles.

E. Acknowledgment

We express our gratitude to Multimedia Nusantara University for the financial assistance and support provided to the research team.

F. References

- [1] K. R. Dörnyei, "Limited edition packaging: objectives, implementations and related marketing mix decisions of a scarcity product tactic," *J. Consum. Mark.*, vol. 37, no. 6, pp. 617–627, 2020.
- [2] A. Stauch, "Does solar power add value to electric vehicles? An investigation of car-buyers' willingness to buy product-bundles in Germany," *Energy Res. Soc. Sci.*, vol. 75, p. 102006, 2021.
- [3] M. Annett, T. Grossman, D. Wigdor, and G. Fitzmaurice, "Exploring and understanding the role of workshop environments in personal fabrication processes," *ACM Trans. Comput. Interact.*, vol. 26, no. 2, pp. 1–43, 2019.
- [4] N. N. Wilim and R. S. Oetama, "Sentiment Analysis About Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, And Decision Tree," *IJNMT (International J. New Media Technol.*, vol. 8, no. 1, pp. 50–56, 2021.
- [5] M. Krishnamoorthy and R. Karthikeyan, "Pattern mining algorithms for data streams using itemset," *Meas. Sensors*, vol. 24, p. 100421, 2022.
- [6] A. Setiawan and F. P. Putri, "Implementasi Algoritma Apriori untuk Rekomendasi Kombinasi Produk Penjualan," *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 66–71, 2020.
- [7] J. Willwersch, *Towards Transparent Real Estate Markets - Assessing*

- Investment Opportunities in Times of Artificial Intelligence*, vol. 100. 2021.
- [8] R. U. Paradkar, "Detecting Customer Purchasing Patterns using Association Rule Mining." Dublin, National College of Ireland, 2022.
 - [9] A. U. Khasanah, "Implementation of market basket analysis based on overall variability of association rule (OCVR) on product marketing strategy," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 722, no. 1, p. 12068.
 - [10] M. A. Christian, N. Nathanael, A. Mauliani, A. Indrawati, L. P. Manik, and Z. Akbar, "Real Market Basket Analysis using Apriori and Frequent Pattern Tree Algorithm," in *ACM International Conference Proceeding Series*, 2021, pp. 161–165. doi: 10.1145/3489088.3489133.
 - [11] F. Bao, L. Mao, Y. Zhu, C. Xiao, and C. Xu, "An Improved Evaluation Methodology for Mining Association Rules," *Axioms*, vol. 11, no. 1, p. 17, 2022, doi: 10.3390/axioms11010017.
 - [12] E. Hikmawati, N. U. Maulidevi, and K. Surendro, "Minimum threshold determination method based on dataset characteristics in association rule mining," *J. Big Data*, vol. 8, no. 1, pp. 1–17, 2021, doi: 10.1186/s40537-021-00538-3.
 - [13] M. D. Ruiz, J. Gómez-Romero, M. Molina-Solana, J. R. Campaña, and M. J. Martin-Bautista, "Meta-association rules for mining interesting associations in multiple datasets," *Appl. Soft Comput. J.*, vol. 49, pp. 212–223, 2016, doi: 10.1016/j.asoc.2016.08.014.
 - [14] J. Hong, R. Tamakloe, and D. Park, "Application of association rules mining algorithm for hazardous materials transportation crashes on expressway," *Accid. Anal. Prev.*, vol. 142, p. 105497, 2020, doi: 10.1016/j.aap.2020.105497.
 - [15] O. El Midaoui, B. El Ghali, A. El Qadi, and R. M. Driss, "Geographical queries reformulation using a parallel association rules generator to build spatial taxonomies," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, pp. 2586–2594, 2021, doi: 10.11591/ijece.v11i3.pp2586-2594.
 - [16] T. M. Alam *et al.*, "A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining," *Biomed. Signal Process. Control*, vol. 68, p. 102726, 2021, doi: 10.1016/j.bspc.2021.102726.
 - [17] F. Z. El Mazouri, M. C. Abounaima, and K. Zenkouar, "Data mining combined to the multicriteria decision analysis for the improvement of road safety: case of France," *J. Big Data*, vol. 6, no. 1, pp. 1–30, 2019, doi: 10.1186/s40537-018-0165-0.