

Analysis of Cigarette Sales Transactions Using Apriori Algorithm at Madura Store

Mochammad Augustiar Mahendra¹, Mamba'us Sa'adah², Erri Wahyu Puspitarini³, Afida Rahman⁴

^{1,2}Department of Informatics Engineering, University of Widya Gama Malang, Jl. Borobudur No. 35 Malang, Indonesia

³Fakultas Teknologi Informasi, Institut Teknologi dan Bisnis Yadika, Pasuruan, Indonesia

⁴Chapai Nawabganj Polytechnic Institute, Chapainawabganj, Bangladesh

Article Info

Article history:

Received March 05, 2025

Revised May 03, 2025

Accepted June 25, 2025

Keywords:

Classification of sales Cigarettes

Apriori Algorithm

Classification model

Data Mining

ABSTRACT

Developments in the cigarette industry continue to increase and there are also challenges in classifying cigarette sales. In this case, the method of classifying cigarette sales using the Apriori algorithm can be one way that can be used. The purpose of this study is to identify significant cigarette sales and classify sales transactions based on sales patterns. The method to be used in this study has several stages. First, we collect cigarette sales data from several different cigarette shops. The data includes information such as transaction ID, items purchased, and sales amounts. Then, we pre-process the data to prepare the raw data for further analysis. The results of this study indicate that classifying cigarette sales using the Apriori algorithm is able to identify significant sales patterns and classify transactions with a more adequate level of accuracy. This research provides new insights in analyzing cigarette sales data and can help decision-making in the cigarette industry.

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

Corresponding Author:

Mochammad Augustiar Mahendra

Department of Informatics Engineering

Faculty of Engineering, University of Widayagama Malang

Jl. Borobudur No. 35 Malang, Jawa Timur

Email: augustiar290821@gmail.com

1. INTRODUCTION

Cigarette sales represent one of the business sectors with significant economic impact in Indonesia. The tobacco industry contributes to national revenue through excise and taxes while providing employment for millions of people [1], [2], [3], [4]. However, to achieve sustainable success, understanding the sales patterns of cigarettes is crucial. The application of the Apriori algorithm provides valuable insights into consumer preferences, sales trends, and the factors influencing cigarette sales. The Apriori algorithm is a popular association rule mining technique that identifies co-occurrence patterns among items [5], [6], [7]. Applying this algorithm in cigarette sales analysis allows producers and retailers to uncover hidden sales patterns in transaction data, thereby improving marketing strategies and inventory management.

The methodology used in this study involves several systematic steps. First, cigarette sales transaction data are gathered from producers, distributors, and retailers, including information on transaction dates, types of cigarettes, quantities sold, sales prices, and other sales-related data. The data undergo preprocessing to clean errors and duplicates and transform it into a binary format suitable for analysis. The Apriori algorithm is then applied to find association rules that meet minimum support and confidence thresholds [8], [9], [10], [11]. For example, a rule may be, "If brand A is sold, then brand B is also sold with a probability of X%." This analysis reveals consumer preferences and relationships between various cigarette products.

Research in Indonesia demonstrates the Apriori algorithm's effectiveness in retail. A study by Herianty et al. applied data mining to analyze consumer purchase patterns in a retail store and showed that association rule mining enhances strategic decision-making [12], [13]. Wulandari and Rahayu used Apriori to redesign store layouts, improving product visibility and increasing sales [14], [15]. These studies highlight how identifying customer purchase patterns can optimize sales and layout strategies. In the cigarette sector, Apriori analysis could

identify surprising purchase behaviors, such as customers who buy brand A also purchasing related items like lighters. This insight enables better product bundling and layout strategies, ultimately increasing sales [15].

Moreover, the Apriori algorithm supports better promotional planning by identifying frequently purchased item combinations. Retailers and manufacturers can create cross-promotions to increase transaction values and customer retention [7]. It also improves inventory management by revealing demand patterns, allowing optimal stock level planning, reducing stockouts, and avoiding excess inventory. Beyond operational advantages, analyzing cigarette sales patterns provides insights into market dynamics, enabling stakeholders to respond more effectively to market changes. In today's digital era, transaction data is a valuable asset for gaining a competitive advantage. Data mining with Apriori helps extract actionable information from complex datasets, empowering cigarette industry players to make data-driven business decisions.

2. METHOD

Customer buying patterns at cigarette shops are often influenced by various factors, such as individual preferences, specific consumption habits, or seasonal trends. Identifying these patterns can not only provide deeper insights into consumer behavior but also serve as the basis for more targeted strategic decision-making.

In this context, the Apriori algorithm was chosen as an analysis tool due to its ability to uncover hidden relationships among items in transaction data. This method has been widely used in various sectors, including retail and services, as a reliable approach for discovering purchasing patterns and relationships between products.

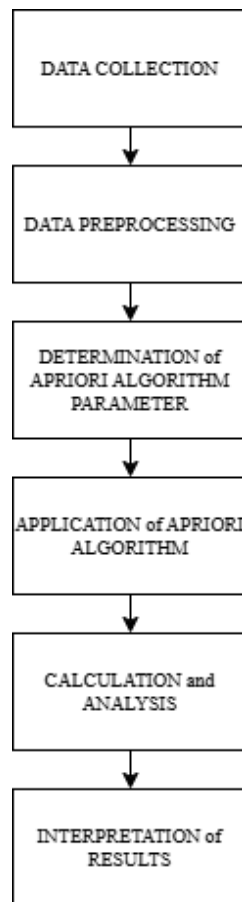


Figure 1. Research Methodology

This research was conducted through several systematic stages to ensure valid and relevant analysis. The stages carried out are as follows:

1. Data Collection:
Collecting cigarette sales transaction data from a shop during a certain period.
2. Data Preprocessing:
Converting transaction data into binary format, where each item is assigned a value of '1' if present in the transaction, and '0' if not.

3. Apriori Algorithm Parameter Determination:
Setting a minimum support value of 30% to determine *frequent itemsets* and a minimum confidence value of 70% to generate strong and relevant association rules.
4. Application of Apriori Algorithm:
Using the Apriori algorithm to find *frequent itemsets* based on the predetermined minimum support value.
5. Calculation and Analysis:
Filtering rules with confidence values above 70% to ensure relevance and analyzing the most significant item combinations.
6. Interpretation of Results:
Developing recommendations based on the findings, such as product bundling strategies or rearranging product layouts in the store.

2.1. Cigarette Sales Transaction Data

The dataset used in this study will be updated to reflect cigarette sales transaction data from a tobacco retailer over a one-year period. Each transaction records various cigarette brands and related products purchased by customers, providing insights into consumer purchasing behaviors and potential associations between different types of cigarettes.

By analyzing this data, we aim to understand which cigarette brands are frequently purchased together within a single transaction. This analysis serves as the foundation for developing association rules that can inform strategies for product bundling, layout optimization, and targeted promotions. The following is a representative example of the cigarette sales transaction data to be analyzed:

Table 1. Cigarette sales transaction data

Transaction	Purchase Items
1	Surya, A Mild, Andalan
2	Surya, A Mild, Malboro
3	Djie sam soe, LA, Malboro
4	JUARA, LA, Surya
5	Djie Sam Soe, A Mild, Malboro
6	JUARA, Surya, A Mild
7	LA, Djie Sam Soe, Diplomat
8	Surya, A Mild, Djie Sam Soe
9	Surya, A Mild, Magnum
10	WIN, Surya, LA
11	WIN, Malboro, Magnum
12	WIN, Surya, Djie Sam Soe

2.2. Frequency Pattern Analysis

In this stage, look for a combination of items that meet the minimum requirements of the support value in the database. The support value in an item is obtained using the following formula:

$$\text{support}(A) = \frac{\text{Number of Transactions Includes } A}{\text{Total Transactions}}$$

Meanwhile, to determine the support value of 2 items, it can be obtained using the following formula:

$$\text{support}(A, B) = \frac{\text{Number of Transactions Includes } A \text{ and } B}{\text{Total Transactions}}$$

Frequently occurring items indicate the set of items that appear more frequently than the predefined minimum limit. After all the frequency patterns have been found, the next step is to find an association rule that fulfils the minimum requirement for confidence by calculating the confidence of the associative rule $A \rightarrow B$. The following is the formula for calculating the confidence value of the rule $A \rightarrow B$:

$$\text{confidence} = P(B | A) = \frac{\text{Number of Transactions Includes } A \text{ and } B}{\text{Total Transactions } A}$$

To determine which association rules to choose, sort them by support x confidence. This rule is obtained based on the number of rules that have the largest value.

2.3. Establishment of Association Rules

To determine which association rule to choose, after finding all the high-frequency samples, the next step is to find the association rule that meets the minimum confidence requirement by calculating the confidence of the association rule. In the case of $A \rightarrow B$, the minimum confidence level is 60%. The value of rule $A \rightarrow B$ is obtained by the following formula

$$\text{confidence} = P(B | A) = \frac{\text{Number of Transactions Includes A and B}}{\text{Total Transactions A}}$$

2.4. Determination of Final Association Rules

At the end of this method, the last step is to determine the final association rules. The final association rules are sorted based on the minimum support and minimum confidence values that have been determined previously.

3. RESULTS AND DISCUSSION

The apriori algorithm is used to match all a priori rules that fulfil the minimum support and minimum confidence requirements. Cigarette sales transaction data was taken from Toko Madura over a period of one year. Total cigarette sales transactions were taken from monthly sales, and then the three months with the highest number of sales were taken to be presented in table 2 as follows:

Table 2. Tabular format of transaction data

Transaksi	Andalan	A Mild	Djie Sam Soe	Diplomat	JUARA	LA	Malboro	Magnum	Surya	WIN
1	1	1	0	0	0	0	0	0	1	0
2	0	1	0	0	0	0	1	0	1	0
3	0	0	1	0	0	1	1	0	0	0
4	0	0	0	0	1	1	0	0	1	0
5	0	1	1	0	0	0	1	0	0	0
6	0	1	0	0	1	0	0	0	1	0
7	0	0	1	1	0	1	0	0	0	0
8	0	1	1	0	0	0	0	0	1	0
9	0	1	0	0	0	0	0	1	1	0
10	0	0	0	0	0	1	0	0	1	1
11	0	0	0	0	0	0	1	1	0	1
12	0	0	1	0	0	0	0	0	1	0

3.1. Analysis of Frequency Patterns for Formation of 1 Itemset

The formation process with 1 itemset or called C1 with the specified minimum amount of support is 30%, as follows:

$$S(\text{And}) = \frac{\text{Transaction Amount Includes And}}{\sum 12} = \frac{1}{12} \times 100\% = 8.33\%$$

$$S(\text{Aml}) = \frac{\text{Transaction Amount Includes Aml})}{\sum 12} = \frac{6}{12} \times 100\% = 50\%$$

$$S(\text{Dss}) = \frac{\text{Transaction Amount Includes Dss}}{\sum 12} = \frac{5}{12} \times 100\% = 41.67\%$$

$$S(\text{Dpt}) = \frac{\text{Transaction Amount Includes Dpt}}{\sum 12} = \frac{1}{12} \times 100\% = 8.33\%$$

$$S(\text{Jua}) = \frac{\text{Transaction Amount Includes Jua}}{\sum 12} = \frac{2}{12} \times 100\% = 16.67\%$$

$$S(LA) = \frac{\text{Transaction Amount Includes LA}}{\Sigma 12} = \frac{4}{12} \times 100\% = 33.33\%$$

$$S(Mlb) = \frac{\text{Transaction Amount Includes Mlb}}{\Sigma 12} = \frac{4}{12} \times 100\% = 33.33\%$$

$$S(Mgm) = \frac{\text{Transaction Amount Includesi Mgm}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16,67\%$$

$$S(Sry) = \frac{\text{Transaction Amount Includes Sry}}{\Sigma 12} = \frac{8}{12} \times 100\% = 66.67\%$$

$$S(WIN) = \frac{\text{Transaction Amount Includes WIN}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16.67\%$$

The support list for each C1 formation item is as follows:

Table 3. Support for each item

Item	Total	Support
Andalan	1	8.33%
A Mild	6	50.00%
Djie Sam	5	41.67%
Soe		
Diplomat	1	8.33%
JUARA	2	16.67%
LA	4	33.33%
Malboro	4	33.33%
Magnum	2	16,67%
Surya	8	66.67%
WIN	2	16.67%

3.2. Formation of 2 itemsets

The next stage is the formation process with 2 itemsets or called C2 with a minimum support of 30%, as follows:

$$S(Amld, Dss) = \frac{\text{Transaction Amount Includes Amld, Dss}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16.67\%$$

$$S(Amld, Sry) = \frac{\text{Transaction Amount Includes Amld, Sry}}{\Sigma 12} = \frac{5}{12} \times 100\% = 41.67\%$$

$$S(Amld, Mlb) = \frac{\text{Transaction Amount Includes Amld, Mlb}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16.67\%$$

$$S(Sry, Dss) = \frac{\text{Transaction Amount Includes Sry, Dss}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16.67\%$$

$$S(Sry, LA) = \frac{\text{Transaction Amount Includes Sry, LA}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16.67\%$$

$$S(Dss, LA) = \frac{\text{Transaction Amount Includes Dss, LA}}{\Sigma 12} = \frac{2}{12} \times 100\% = 16.67\%$$

$$S(Dss, Mlb) = \frac{\text{Transaction Amount Includes Dss, Mlb}}{\sum 12} = \frac{2}{12} \times 100\% = 16.67\%$$

The support list of each C2 formation item is:

Table 4. Minimum results support 2 items

Item	Total	Support
A Mild, Djie Sam Soe	2	16.67%
A Mild, Surya	5	41.67%
A Mild, Malboro	2	16.67%
Surya, Djie Sam Soe	2	16.67%
Surya, LA	2	16.67%
Djie Sam Soe, LA	2	16.67%
Djie Sam Soe, Malboro	2	16.67%

From the process of forming itemsets in table 4 with a minimum support of 30%, it is known that the combination of 2 itemsets A Mild, Surya with a support of 41.67% is on cigarettes A Mild, Surya.

Because the formation of 3 itemsets is not obtained at 30%, the process of making itemsets is only up to 2 combinations.

3.3. Establishment of Association Rules

From the combination of the 2 itemsets formed, it can be observed that the confidence value is:

Table 5. Pembentukan aturan asosiasi

Rule	Confidence	
If you buy A Mild then buy Surya	5/6	83.33%
If you buy Surya then buy A Mild	5/8	62.5%

3.4. Final Association Rules

The final association rules are ordered based on the minimum support and confidence that have been determined, as follows:

Table 6. Establishment of final association rules

Rule	Support	Confidence
If you buy A Mild then buy Surya	41.67%	83.33%
If you buy Surya then buy A Mild	41.67%	62.5%

Based on the data in table 6, it can be seen that the cigarettes most frequently purchased by consumers are A Mild and Surya with a support value of 41.67% and a confidence value of 83.33%.

The rule means '83.33% of the overall transactions containing the A Mild item also contain the Surya item, while 41.67% of all existing transactions contain both items. This means that, a consumer who buys A Mild has an 83.33% probability of buying Surya. This rule is quite significant as it represents 50% of the transaction records so far.

Meanwhile, Surya and A Mild have a support value of 41.67% and a confidence value of 62.5%. The rule means '62.5% of all transactions containing Surya items also contain A Mild items, while 41.67% of all transactions contain both items. That means a consumer who buys Surya has a 62.5% chance of buying A Mild. This rule is quite significant as it represents 50% of the transaction records so far.

By knowing the cigarettes that are most often bought by consumers, the store can develop a strategy in determining cigarette purchases to maintain the availability of stock needed by consumers and can also arrange the layout of cigarettes based on the combination of cigarette itemsets formed.

4. CONCLUSION

In the context of cigarette sales, the use of data mining and Apriori algorithm can be very useful to find out cigarette sales patterns. In the context of cigarette sales, the use of this algorithm can help cigarette manufacturers or traders to identify sales patterns that may be unknown in the transaction data, so as to make better decisions in marketing and inventory management.

It is expected that the results of the analysis of cigarette sales patterns using the Apriori algorithm will provide valuable insights for cigarette manufacturers, distributors, and retailers to optimise marketing strategies, inventory management, and better decision-making in cigarette sales.

REFERENCES

- [1] S. M. A. Ade Irma Amanda, Debi Setiawan, and Liza Trisnawati, “Penerapan Algoritma Apriori Dalam Menganalisis Pola Minat Beli Konsumen Di Coffee Shop,” *JEKIN - Jurnal Teknik Informatika*, vol. 3, no. 1, pp. 25–32, Jul. 2023, doi: 10.58794/jekin.v3i1.483.
- [2] A. Christian, R. Supriyadi, F. Ariani, and Amir, “IMPLEMENTASI ALGORITMA APRIORI DALAM MENENTUKAN POLA PEMBELIAN PADA TOKO UMBAH SEPATU,” *Jurnal Informatika dan Rekayasa Elektronik*, vol. 6, no. 2, pp. 162–169, Nov. 2023, doi: 10.36595/jire.v6i2.840.
- [3] N. Hadinata and K. Kurniawan, “ANALISIS POLA PEMBELIAN PRODUK MAKANAN RINGAN MENGGUNAKAN ALGORITMA APRIORI,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 1, pp. 1–7, Feb. 2020, doi: 10.32736/sisfokom.v9i1.623.
- [4] V. S. Rifania, S. Saniman, and A. Azlan, “Penerapan Algoritma Apriori Dalam Mencari Pola Pembelian Konsumen,” *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 2, p. 201, Mar. 2023, doi: 10.53513/jursi.v2i2.5750.
- [5] T. Prasetya, J. E. Yanti, A. I. Purnamasari, A. R. Dikananda, and O. Nurdiawan, “Analisis Data Transaksi Terhadap Pola Pembelian Konsumen Menggunakan Metode Algoritma Apriori,” *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, vol. 6, no. 1, p. 43, Apr. 2022, doi: 10.51211/itbi.v6i1.1688.
- [6] S. Diwandari and U. Zaky, “Analysis of Customer Purchase Behavior using Association Rules in e-Shop,” in *2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, Nov. 2021, pp. 144–149. doi: 10.1109/ICITISEE53823.2021.9655892.
- [7] E. H. Hermaliani, L. Kurniawati, T. Haryanti, N. Mutiah, A. Kurniawan, and B. S. Renhoran, “Data Mining Technique to Determine the Pattern of Fruits Sales & Supplies Using Apriori Algorithm,” *J Phys Conf Ser*, vol. 1641, no. 1, p. 012070, Nov. 2020, doi: 10.1088/1742-6596/1641/1/012070.
- [8] X. Zhang and J. Zhang, “Analysis and research on library user behavior based on apriori algorithm,” *Measurement: Sensors*, vol. 27, p. 100802, Jun. 2023, doi: 10.1016/j.measen.2023.100802.
- [9] D. T. Tran and J.-H. Huh, “Building a model to exploit association rules and analyze purchasing behavior based on rough set theory,” *J Supercomput*, vol. 78, no. 8, pp. 11051–11091, May 2022, doi: 10.1007/s11227-021-04275-5.
- [10] A. Sinha, “Implying Association Rule Mining and Market Basket Analysis for Knowing Consumer Behavior and Buying Pattern in Lockdown - A Data Mining Approach,” May 06, 2021. doi: 10.20944/preprints202105.0102.v1.
- [11] A. A. Aldino, E. D. Pratiwi, Setiawansyah, S. Sintaro, and A. Dwi Putra, “Comparison Of Market Basket Analysis To Determine Consumer Purchasing Patterns Using Fp-Growth And Apriori Algorithm,” in *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, IEEE, Oct. 2021, pp. 29–34. doi: 10.1109/ICOMITEE53461.2021.9650317.
- [12] M. H. Santoso, “Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom,” *Brilliance: Research of Artificial Intelligence*, vol. 1, no. 2, pp. 54–66, Dec. 2021, doi: 10.47709/brilliance.v1i2.1228.
- [13] R. I. Pramesti and E. Gravitiani, “IMPACT OF TOBACCO EXCISE ON SECTORAL OUTPUT AND GROSS INCOME OF INDONESIA : AN APPLICATION OF INPUT OUTPUT ANALYSIS,” *Journal of Finance, Economics and Business*, vol. 1, no. 2, pp. 83–108, May 2023, doi: 10.59827/jfeb.v1i2.68.
- [14] M. H. Santoso, “Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom,” *Brilliance: Research of Artificial Intelligence*, vol. 1, no. 2, pp. 54–66, Dec. 2021, doi: 10.47709/brilliance.v1i2.1228.
- [15] A. R. Efrat, R. Gernowo, and Farikhin, “Consumer purchase patterns based on market basket analysis using apriori algorithms,” *J Phys Conf Ser*, vol. 1524, no. 1, p. 012109, Apr. 2020, doi: 10.1088/1742-6596/1524/1/012109.