



Utilization of Sales Data Analysis for Product Recommendation Systems in E-Commerce Using the Apriori Algorithm

Muhammad Noor Hasan Siregar^{1,*}, Furqan Khalidy², Rismayanti³, Khairunnisa⁴

¹Fakultas Ekonomi, Bisnis Digital, Universitas Graha Nusantara, Indonesia

²Prodi Ilmu Komputer, Universitas Nahdlatul Ulama Sumatera Utara, Indonesia

³Prodi Teknik Informatika, Univ Harapan Medan, Sumatera Utara, Indonesia

⁴Prodi Teknik Informatika, Universitas Harapan Medan, Sumatera Utara, Indonesia

Author(s) Email: ¹noor.siregar@gmail.com, ²furqankhalidy@gmail.com, ³risma.stth@gmail.com, ⁴khairunnisajv2@gmail.com

ARTICLE INFO

Article history:

Received July 10, 2024

Revised July 12, 2024

Accepted July 12, 2024

Publish November 30, 2024

ABSTRACT

The rapid development of e-commerce has significantly increased the volume of sales transactions and customer interaction data. This presents an opportunity for businesses to leverage data mining techniques to extract valuable insights that support decision-making processes. One such application is the development of product recommendation systems, which play a crucial role in enhancing customer satisfaction and driving sales. This research focuses on utilizing sales transaction data to build a product recommendation system using the Apriori algorithm, a well-known method for association rule mining. The study begins with the collection and preprocessing of transaction data from an e-commerce platform. Through the application of the Apriori algorithm, frequent itemsets are identified, and association rules are generated based on specified support and confidence thresholds. These rules reveal purchasing patterns and relationships between products that are frequently bought together. The system then uses these patterns to recommend relevant products to users, aiming to improve cross-selling opportunities and personalize the shopping experience. The results demonstrate that the Apriori-based recommendation model is effective in identifying meaningful product combinations and can be implemented as a lightweight, interpretable alternative to more complex machine learning methods. Furthermore, the system helps e-commerce businesses optimize inventory management and marketing strategies by understanding customer buying behavior. This research concludes that the integration of the Apriori algorithm into recommendation systems provides tangible benefits for e-commerce platforms seeking data-driven personalization solutions.

Keywords:

Data Mining, Apriori Algorithm, E-Commerce, Recommendation System, Sales Data Analysis

Corresponding Author:

Muhammad Noor Hasan Siregar,

Fakultas Ekonomi, Bisnis Digital, Universitas Graha Nusantara, Indonesia

Email: noor.siregar@gmail.com



1. INTRODUCTION

The development of information and communication technology has had a significant impact on various sectors of life, one of which is the trade sector[1]. E-commerce or electronic commerce has now become one of the fastest-growing business models in the digital era[2]. This phenomenon is closely related to the ease of internet access, changes in consumer behavior, and the increasing use of digital devices in daily activities[3]. E-commerce offers various conveniences for consumers, such as purchasing goods online without having to visit physical stores, digital payment systems, and a wide range of products[4]. For business operators, e-commerce opens up significant opportunities to reach a broader market, improve operational efficiency, and collect consumer data in real-time[5]. One of the challenges faced by e-commerce practitioners is how to provide a personalized and relevant shopping experience for each user. In this context, recommendation systems play an important role[6]. Recommendation systems are technologies used to provide product suggestions or recommendations to users based on their preferences, behavior, or interaction history on the e-commerce platform. This system has proven capable of increasing sales, enhancing customer loyalty, and helping consumers find products they might need but were not previously aware of[7].

To build an effective recommendation system, an in-depth analysis of the available data is needed. One type of data that is very important in the context of e-commerce is sales transaction data[8]. This data includes information about which products are purchased by customers, when the purchases are made, and the purchasing patterns that occur between products. Through the analysis of transaction data, e-commerce practitioners can identify products that are often bought together, customer shopping patterns, and emerging market trends. In this research, the Apriori algorithm is used as the main method to analyze sales data. The Apriori algorithm is one of the most popular algorithms in data mining, particularly in the field of association rule mining[9]. The main objective of this algorithm is to find associative patterns between items that frequently appear together in a transaction dataset. In other words, this algorithm can be used to determine which products are likely to be purchased together by customers. The result of this algorithm consists of frequent itemsets and association rules that can be utilized to build an association-based recommendation system[10].

The reason for choosing the Apriori algorithm in this research is because this algorithm has the advantage of result interpretation that is easy to understand for non-technical business users. Additionally, this algorithm works efficiently on datasets with a large number of transactions and is capable of capturing logical relationships between products in the form of clear rules. Although there are various algorithms and other approaches in building recommendation systems, such as collaborative filtering, content-based filtering, and more complex machine learning-based techniques, the Apriori-based approach remains relevant and highly valuable, especially for needs that emphasize readability and transparency of results[11]. This research aims to utilize the Apriori algorithm in processing sales transaction data to generate a relevant product recommendation system on e-commerce platforms[12]. The focus of the research includes the process of collecting and preprocessing transaction data, applying the Apriori algorithm with specific support and confidence parameters, and analyzing the results in the form of discovered association rules. In addition, this research also evaluates the effectiveness of the built recommendation system in the context of increasing the relevance of product suggestions and the potential for sales growth[13].

With the presence of a recommendation system based on sales data analysis, e-commerce players can provide product suggestions that are more accurate and aligned with consumer shopping patterns[14]. This will certainly contribute to the improvement of user experience, customer loyalty, and the effectiveness of marketing strategies[15]. On the other hand, this approach also opens up new opportunities in the development of data-driven sales strategies, such as product bundling arrangements, product placement on the main page, and promotions for products that are strongly related to each other.

This research is expected to provide a significant contribution both academically and practically. From an academic perspective, this research strengthens the application of data mining theory in a real-world context, namely e-commerce, while also enriching the reference for case studies on the use of the Apriori algorithm for recommendation systems. Meanwhile, from a practical standpoint, the results of this research can serve as a guide or model for business practitioners in developing their internal recommendation systems by utilizing already available data without requiring expensive or complex technology. Thus, the use of the Apriori algorithm for sales data analysis can become an efficient, accurate, and easily applicable strategic solution for various scales of e-commerce businesses.

2. RESEARCH METHODOLOGY

This research uses a quantitative approach with an exploratory method, aimed at analyzing sales transaction data and building a product recommendation system based on the Apriori algorithm. This method was chosen because it is suitable for uncovering hidden patterns in the data and finding association rules between items in the transaction dataset. The research was conducted in stages through several phases as follows.



Figure 1. Research Flow

2.1 Data Collection

The data used in this research consists of sales transaction data from an e-commerce platform. This data contains customer transaction records that include transaction ID, product ID, transaction time, and customer information in an anonymous manner. Data can be obtained from the relevant e-commerce backend system or by using public datasets as a simulation, such as the Online Retail Dataset or Groceries Dataset. The minimum number of transaction data should be in the hundreds of rows to ensure diversity and analytical strength.

2.2 Data Preprocessing

Before being applied to the Apriori algorithm, the collected transaction data needs to undergo a preprocessing process, including:

- Data cleaning: Removing duplicates, empty data, or invalid transactions.
- Data transformation: Converting transaction data into a basket format suitable for associative analysis, where each row represents a single transaction and contains a collection of products purchased together.
- Data normalization: If necessary, standardize the format of product names or product IDs to avoid analysis errors.

2.3 Application of the Apriori Algorithm

After the data is ready, the analysis process is carried out using the Apriori algorithm with the following steps:

- Determining the minimum support and confidence values to be used as algorithm parameters. These values are adjusted based on the characteristics of the dataset, for example, support 0.05 (5%) and confidence 0.6 (60%).
- Running the algorithm to obtain frequent itemsets, which are combinations of products that are often purchased together.
- Generating association rules, which are rules in the form of "If buying A, then likely also buying B."
- Evaluating the rules formed based on lift values to determine the strength of the correlation between products.

2.4 Development of the Recommendation System

The results of the Apriori algorithm in the form of association rules are utilized to build a product recommendation system. This system will recommend products to users based on the products they have seen or purchased. The basic logic of the system is:

- If the user buys product A, the system will recommend product B, which has a strong association with A.
- The system was developed in the form of a simple web-based prototype or Python script as a proof of concept.

2.5 System Evaluation

Evaluation is conducted to measure the quality of the recommendations produced by the system. This evaluation includes:

- Quantitative evaluation, which involves counting the number of valid and relevant rules based on lift, support, and confidence values.
- Qualitative evaluation, which involves manually testing the recommendation results by comparing them with domain knowledge or simulated customer preferences.
- User feedback (optional): if the research involves users, feedback related to the relevance of the product suggestions that emerge can be collected.

2.6 Tools and Software

In this research, several devices and supporting libraries were used, including:

- Python programming language
- MLxtend or apyori library for implementing the Apriori algorithm
- Microsoft Excel or Google Sheets for initial data exploration
- Jupyter Notebook as a medium for development and visualization

3. RESULT AND DISCUSSION

After conducting the sales data analysis process using the Apriori algorithm, a number of frequent itemsets and association rules were obtained, illustrating customer purchasing patterns on the e-commerce platform. This process was carried out using a transaction dataset of 1,000 entries, which had undergone pre-processing and transformation into basket format. The parameters used in the Apriori algorithm are a minimum support of 5% and a minimum confidence of 60%.

Here is an example of the association rules obtained:

Table 1. Example of Association Rules Results

No	Association Rules	Support (%)	Confidence (%)	Lift
1	If you buy a Keyboard → Mouse	6.8	75.0	1.92
2	If you buy a laptop → Laptop bag	5.5	72.3	1.87
3	If you buy a printer → Ink	6.2	80.1	2.13
4	If you buy a Mouse → Mousepad	7.3	69.4	1.65
5	If you buy a smartphone → Casing	5.9	74.2	1.81

Explanation:

- a. Support shows how often the combination of items appears in the overall transactions.
- b. Confidence shows the probability of a consumer buying item B if they have already bought item A.
- c. Lift describes the strength of the association between two items; a value above 1 indicates a positive correlation.

3.1 Analysis of Results

From the results above, it can be seen that there is a fairly strong purchasing pattern between products. For example:

- a. Rule number 3, which states that if you buy a printer, you are very likely to also buy ink, has the highest confidence at 80.1% and a lift of 2.13, indicating a very strong association. This aligns with business logic that customers who buy printers usually need ink as a complement.
- b. Rule number 1 and 5 also show a high correlation, where Keyboard and Mouse, as well as Smartphone and Casing, are reasonable product pairs that are often purchased together.

This result is very useful in the development of product recommendation systems. For example, if a user adds a Printer to their cart, the system can automatically recommend Ink as a relevant product. This recommendation not only enhances user convenience but also opens up opportunities for additional purchases (cross-selling), thereby increasing revenue for e-commerce operators.

Furthermore, the high lift values in most rules indicate that the associations between products in the dataset are quite strong and can serve as a basis for marketing strategies, such as product bundling arrangements, combination discounts, or targeted promotions.

3.2 Visualization of Association Rules (Optional)

If used in presentations or publications, these association results can also be visualized in the form of graphs or network maps to show the relationships between products interactively. This visualization will make it easier for the business team to understand the interconnections between items.

3.3 Discussion of Advantages and Disadvantages

- a. Advantages of the Apriori method:
 - 1 The results are easy to understand and interpret by non-technical users.
 - 2 Suitable for initial data exploration before applying more complex methods.
 - 3 Can be adjusted flexibly based on support and confidence parameters.
- b. Shortcomings:
 - 1 If the dataset is very large, the computation process can take a lot of time and resources.
 - 2 Not considering the order of time or the context of the transaction.
 - 3 It is necessary to set the correct threshold to avoid generating too many irrelevant rules..

3.4 Practical Implications

The results of this analysis can be directly utilized in the product recommendation system on e-commerce platforms. By utilizing this association rule:

- a. Product recommendations can be provided in real-time when customers add products to their cart.
- b. Can be used for upselling and cross-selling.
- c. Making it easier for product managers to create promotional packages or bundling discounts:

4. CONCLUSIONS

This research shows that the utilization of sales transaction data in e-commerce can serve as a strong foundation for building an effective and relevant product recommendation system. By using the Apriori algorithm, purchasing patterns that frequently occur among products can be identified in the form of association rules that are easy to understand and apply. These rules can be used to provide automatic product recommendations to customers, which can ultimately enhance user comfort, speed up the shopping process, and encourage additional purchases (cross-selling and upselling). The analysis results show that several product pairs have high support and confidence levels, as well as lift values above 1, indicating a strong correlation between those products. For example, the relationship between the purchase of printers

and ink, or laptops and laptop bags, is a common pattern that can be utilized in recommendation systems to increase sales opportunities. From a technical standpoint, the Apriori algorithm offers advantages in terms of result transparency and flexibility in parameter settings. However, it is important to note that the use of this algorithm also has limitations, especially when applied to very large or complex datasets. Therefore, in its implementation, it is necessary to consider computational efficiency and periodic result validation. Overall, this research proves that sales data analysis using the Apriori algorithm can provide strategic value in the development of data-driven recommendation systems on e-commerce platforms. This approach can serve as an initial solution that is light, practical, and beneficial for business operators in providing more personalized services to customers..

REFERENCES

- [1] A. A. Nabi, F. H. Tunio, M. Azhar, M. S. Syed, and Z. Ullah, "Impact of information and communication technology, financial development, and trade on economic growth: Empirical analysis on N11 countries," *J. Knowl. Econ.*, vol. 14, no. 3, pp. 3203–3220, 2023.
- [2] A. Sharma, S. K. Mishra, and V. K. Srivastav, "The Evolution And Impact Of E-Commerce.," *J. Namibian Stud.*, vol. 33, 2023.
- [3] E. E. Pramiasih, "Consumer behavior in the digital era," *Int. J. Financ. Econ.*, vol. 1, no. 3, pp. 662–674, 2024.
- [4] Z. Kedah, "Use of e-commerce in the world of business," *Startupreneur Bus. Digit. (SABDA Journal)*, vol. 2, no. 1, pp. 51–60, 2023.
- [5] O. Ogunwolu, E. C. Onukwulu, N. J. Sam-Bulya, M. O. Joel, and G. O. Achumie, "Optimizing automated pipelines for realtime data processing in digital media and e-commerce," *Int. J. Multidiscip. Res. Growth Eval.*, vol. 3, no. 1, pp. 112–120, 2022.
- [6] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. M. Alam, and S. Luo, "A review of content-based and context-based recommendation systems," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 3, pp. 274–306, 2021.
- [7] A. A. Agha, A. Rashid, R. Rasheed, S. Khan, and U. Khan, "Antecedents of customer loyalty at telecomm sector.," *Turkish Online J. Qual. Inq.*, vol. 12, no. 9, 2021.
- [8] N. Johnson, M. S. Purwanegara, and N. B. Mulyono, "Enhancing E-commerce with big data: from browsing to buying through recommendation systems," *Int. J. Entrep. Bus. Creat. Econ.*, vol. 4, no. 1, p. 130, 2024.
- [9] S. Khodijah, C. A. Rizki, and M. Hasanuddin, "Journal of Computer Science Artificial Intelligence," vol. 1, no. 1, pp. 1–6, 2024.
- [10] M. Hasanuddin, "Journal of Computer Science Artificial Intelligence Optimization of Computer Network Performance Using Heuristic Algorithms," vol. 1, no. 1, pp. 12–17, 2024.
- [11] L. Valtonen, S. J. Mäkinen, and J. Kirjavainen, "Advancing reproducibility and accountability of unsupervised machine learning in text mining: Importance of transparency in reporting preprocessing and algorithm selection," *Organ. Res. Methods*, vol. 27, no. 1, pp. 88–113, 2024.
- [12] A. Kurniawan and N. Suwaryo, "Analysis of the Apriori Algorithm for Enhancing Retail Product Staple Sales Recommendations," *Int. J. Softw. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 449–456, 2023.
- [13] M. Hasanuddin, B. E. Susanto, S. Ginting, and F. Rizaldi, "Analisis Minat Siswa Kelas 1 SMK Pada Ekstrakurikuler Sepak Bola Dengan Metode Technology Acceptance Model," vol. 4, no. 1, pp. 52–58, 2025.
- [14] F. T. A. Hussien, A. M. S. Rahma, and H. B. A. Wahab, "Recommendation systems for e-commerce systems an overview," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 12024.
- [15] A. T. Reni, "The Influence Of Personalization, User Experience, And Digital Marketing Strategy On Customer Loyalty With Churn Rate As A Mediating Variable In The E-Commerce Industry," *J. Ilmu Sos. Mamangan*, vol. 12, no. 3, pp. 1522–1530, 2024.