

Analysis of Consumer Purchase Patterns on Google Merchandise Store E-Commerce Using Apriori Algorithm to Support Marketing Strategies

Qaiz Ramzy Julianto¹, Andika Prayoga², Ghifari Akbar³, Jeffersan Christian⁴, Anita Ade Rahma⁵

^{1,2,3,4,5}Universitas Putra Indonesia YPTK, Padang, Indonesia

¹qaizramzy@gmail.com, ²andikaprayoga7679@gmail.com, ³ghifariakbar.fari01@gmail.com, ⁴jeffersansans@gmail.com,

⁵anita_aderahma@upiypk.ac.id

Article Info

Article history:

Received February 18, 2026

Revised February 19, 2026

Accepted February 19, 2026

Keywords:

Data mining

Association rule

Apriori

E-commerce

Market basket analysis

ABSTRACT

The development of e-commerce generates large amounts of transaction data that can potentially be used to support data-driven marketing strategies. This study aims to analyze consumer purchase patterns in the Google Merchandise Store dataset using the Association Rule Mining method with apriori algorithm. The dataset used consisted of 12,267 transactions with 382 product attributes transformed into a binary incidence matrix. The parameters applied are a minimum support of 0.005 and a minimum confidence of 0.05. The results of the analysis resulted in seven association rules that met the threshold. The rule with the highest confidence shows that 47.1% of consumers who buy Android Sticker Sheet Ultra Removable also buy Google Laptop and Cell Phone Stickers. These findings indicate the existence of product linkages that can be used for cross-selling and product bundling strategies. Overall, the apriori algorithm has proven to be effective in identifying consumer purchasing patterns as well as supporting data-driven marketing decision-making on e-commerce platforms.

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



Corresponding Author:

Anita Ade Rahma

1. INTRODUCTION

The development of information technology has driven the growth of e-commerce significantly in recent years [1]. Digital platforms generate large amounts of transaction data that record consumer purchasing behavior in detail [2]. The data contains important information about the pattern of interconnectedness between products purchased at the same time. However, in practice, many e-commerce players have not made optimal use of transaction data to support strategic decision-making, such as determining product bundling, promotion strategies, and recommendation systems [3].

The main problem in managing e-commerce transaction data is how to systematically identify consumer purchase patterns from datasets that have a high number of transactions and product variations [4]. Without a proper analytical approach, transaction data only becomes a historical archive that does not provide added value for the development of marketing strategies [5]. Therefore, a method that is able to extract hidden relationships between products is needed based on the frequency of occurrence and the degree of their relevance in transactions.

In the field of data mining, Association Rule Mining is one of the techniques used to find patterns of relationships between items in transaction datasets [6]. Apriori algorithms are known to be effective in identifying frequent itemsets and forming association rules based on support and confidence parameters [7].

Various previous studies have shown that the application of apriori to retail and e-commerce data is able to support market basket analysis and increase the effectiveness of promotional strategies and product recommendations [8]. However, implementations on datasets with high product variety still require parameter adjustments to produce relevant and non-redundant rules.

This study proposes the application of apriori algorithm to the Google Merchandise Store transaction dataset to identify consumer purchase patterns. The analysis process is carried out through the stages of data preprocessing in the form of a binary incidence matrix, determining minimum support and minimum confidence, and evaluating the rules of association formed. This approach aims to produce rules that are not only statistically significant, but also have practical implications in e-commerce marketing strategies.

The novelty value of this study lies in the application of apriori algorithm on the Google Merchandise Store dataset with a relatively large number of product attributes, as well as the analysis of product linkages based on exploratively adjusted parameters. This research not only displays association rules, but also interprets the results in the context of digital business strategies, thus making a theoretical contribution to the development of association rule mining and a practical contribution to data-driven marketing decision-making.

2. METHOD

2.1. Research Design

This study uses a quantitative approach with data mining methods, especially Association Rule Mining using apriori algorithm. The research design is exploratory because it aims to identify patterns of interconnectedness between products in e-commerce transaction data. The research was carried out systematically according to the stages shown in the Research Flow Diagram, which consisted of the stages of data collection, preprocessing, application of apriori algorithms, analysis of association rules, to interpretation of results and drawing conclusions.

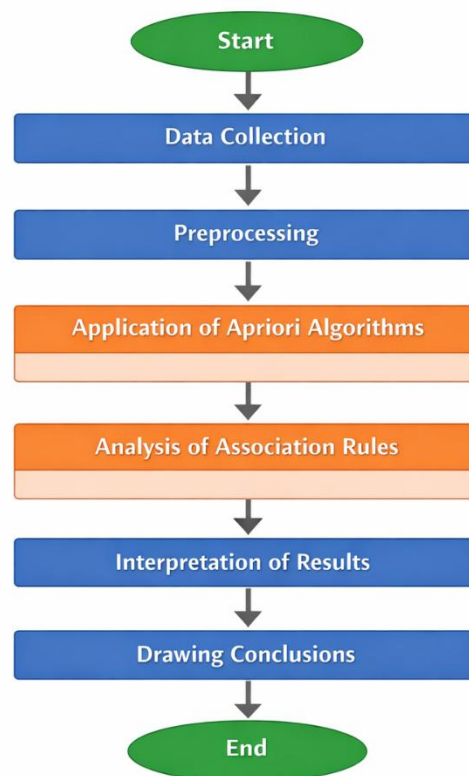


Figure 1. Research Flow Diagram

2.2. Research Stages

2.2.1. Data Collection

The initial stage of the research is the collection of transaction data from the Google Merchandise Store dataset obtained through the Kaggle platform. The dataset used consisted of 12,267 transactions with 382 product attributes.

Table 1. Transaction Fata from the Google Merchandise Store

transaction_id	Google Zip Hoodie Black	Google Women's Short Sleeve Hero Tee White	Google Trike Tee Black	...	Google Women's Short Sleeve Badge Tee Grey
16681	FALSE	FALSE	FALSE	...	FALSE
16682	FALSE	FALSE	FALSE	...	FALSE
16685	FALSE	TRUE	FALSE	...	FALSE
16687	FALSE	FALSE	FALSE	...	FALSE
16692	FALSE	FALSE	FALSE	...	FALSE
16693	FALSE	FALSE	FALSE	...	FALSE
16695	FALSE	FALSE	FALSE	...	FALSE
16696	FALSE	FALSE	FALSE	...	FALSE
16697	FALSE	FALSE	FALSE	...	FALSE
16704	FALSE	FALSE	FALSE	...	FALSE
...	...	...	...	...	...
48494	TRUE	FALSE	FALSE	...	FALSE

Table 1 presents examples of transaction data that contain several products, such as Google Zip Hoodie Black, Google Women's Short Sleeve Hero Tee White, Google Trike Tee Black, and Google Women's Short Sleeve Badge Tee Grey. A value of TRUE indicates that the product was purchased in a transaction, while a value of FALSE indicates that there were no purchases on the product. The data displayed is secondary data in the form of consumer purchase transaction history on e-commerce platforms that has been transformed into binary format for analysis purposes.

2.2.2. Preprocessing Data

At this stage, the process of cleaning and transforming the data is carried out to adjust to the needs of the algorithm apriori. The preprocessing step includes the selection of relevant product attributes, the transformation of transaction data into the form of a binary incidence matrix, and the representation of data in binomial form (true/false) or 1 and 0. This transformation aims to simplify the process of searching for frequent itemset and improve computing efficiency.

2.2.3. Application of Apriori Algorithm

After the data is ready to be analyzed, an apriori algorithm is applied using RapidMiner software. The parameters used in this study were a minimum support value of 0.005, and a minimum confidence value of 0.05. The determination of minimum support and minimum confidence was carried out by considering the characteristics of the dataset consisting of 12,267 transactions and 382 product attributes with high item variation. The minimum support of 0.005 (0.5%) is equivalent to about 61 transactions and is chosen to filter out fairly representative product combinations and avoid noise. In the context of e-commerce, support below 1% is still relevant to identify meaningful patterns [6]. A minimum confidence of 0.05 (5%) is used to maintain a balance between the number of rules and potential association patterns. Parameters are determined through a trial and error approach so that the resulting rules are optimal and remain analytically relevant. The apriori algorithm works by identifying the frequent itemset based on the minimum support value, then forming association rules based on the confidence value [9].

2.2.4. Analysis Association Rules

In this study, the formation of association rules uses two main parameters, namely support and confidence. Support is used to measure the rate at which a combination of items appears in an overall transaction [10], while confidence is used to measure the level of confidence or probability of a consequential item occurring when an antecedent item appears [11]. Mathematically, the support value for the $A \rightarrow B$ association rule is formulated based on equation 1.



$$\text{Support } A \rightarrow B = \frac{\sigma(A \cap B)}{N} \quad (1)$$

Where $\sigma(A \cap B)$ expresses the number of transactions containing items A and B simultaneously, and N is the total number of transactions in the dataset. Meanwhile, the confidence value is formulated based on equation 2.

$$\text{Confidence } A \rightarrow B = \frac{\sigma(A \cap B)}{\sigma(A)} \quad (2)$$

Where $\sigma(A)$ is the number of transactions that contain item A. Confidence value indicates the probability of consumers buying product B on the condition that they have purchased product A. The higher the confidence value, the stronger the association relationship between products that is formed [12]. Through the calculation of these two parameters, the Apriori algorithm is able to identify the frequent itemset as well as generate relevant association rules to be further analyzed in the context of e-commerce marketing strategies [13]. Thus, a rule with a high confidence value is interpreted as a product relationship that has a strong tendency to be purchased simultaneously [14].

2.2.5. Interpretation of Results and Evaluation

The results of association rules are interpreted in the context of e-commerce marketing strategies, such as cross-selling, product bundling, and product recommendations. The evaluation is carried out by assessing the relevance of the rules to consumer purchasing behavior and their potential implementation in the recommendation system.

2.2.6. Conclusion Drawdown

The final stage of the research is to formulate conclusions based on the results of the association rules analysis and provide implications and suggestions for further research development.

3. RESULTS AND DISCUSSION

Based on transaction data processing using apriori algorithm through the RapidMiner software, seven association rules were obtained that met the minimum support parameters of 0.005 and the minimum confidence of 0.05. The analyzed dataset consisted of 12,267 transactions with 382 product attributes in the form of a binary incidence matrix.

AssociationRules

```
Association Rules
[Android Sticker Sheet Ultra Removable] --> [8 pc Android Sticker Sheet] (confidence: 0.316)
[YouTube Custom Decals] --> [Google Laptop and Cell Phone Stickers] (confidence: 0.319)
[8 pc Android Sticker Sheet] --> [Google Laptop and Cell Phone Stickers] (confidence: 0.349)
[DV] --> [Google Laptop and Cell Phone Stickers] (confidence: 0.384)
[Keyboard DOT Sticker] --> [Google Laptop and Cell Phone Stickers] (confidence: 0.388)
[HG] --> [Google 22 oz Water Bottle] (confidence: 0.468)
[Android Sticker Sheet Ultra Removable] --> [Google Laptop and Cell Phone Stickers] (confidence: 0.471)
```

Figure 2. Results of Association Rules

The resulting association rules show that there is a connection between purchases between products in Google Merchandise Store transactions. Some rules with relatively high confidence values include:

Android Sticker Sheet Ultra Removable → Google Laptop and Cell Phone Stickers (confidence: 0,471)

HG → Google 22 oz Water Bottle (confidence: 0,468)

Keyboard DOT Sticker → Google Laptop and Cell Phone Stickers (confidence: 0,388)

DV → Google Laptop and Cell Phone Stickers (confidence: 0,384)

8 pc Android Sticker Sheet → Google Laptop and Cell Phone Stickers (confidence: 0,349)

YouTube Custom Decals → Google Laptop and Cell Phone Stickers (confidence: 0,319)

Android Sticker Sheet Ultra Removable → 8 pc Android Sticker Sheet (confidence: 0,316)

The rule with the highest confidence value is Android Sticker Sheet Ultra Removable → Google Laptop and Cell Phone Stickers of 0.471. This shows that 47.1% of transactions containing Android Sticker Sheet Ultra Removable products also included Google Laptop and Cell Phone Stickers products in the same transaction.

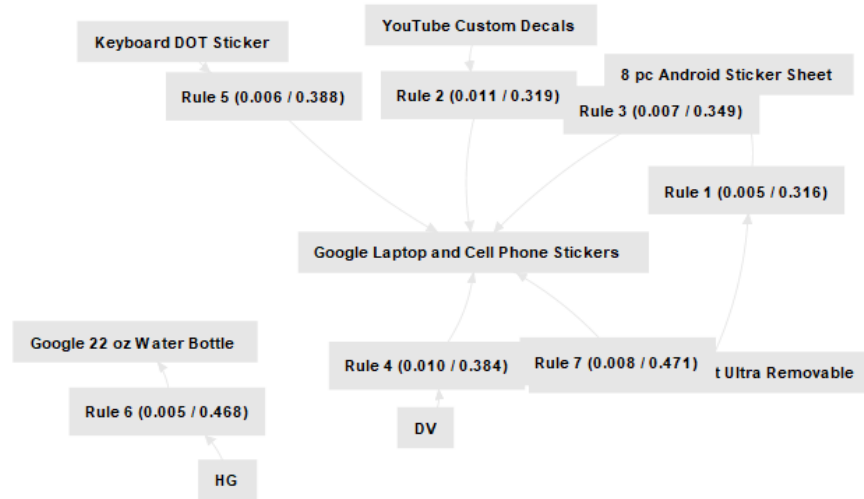


Figure 3. Graph of Relationships Between Products

Visualizations of the relationships between products are displayed in the form of association graphs, where Google Laptop and Cell Phone Stickers appear as central nodes that have a lot of interconnectedness with other products. This shows that these products have a significant role in consumer purchasing patterns. The results of the study show that the apriori algorithm is able to effectively identify consumer purchase patterns on e-commerce datasets with a large number of attributes. The formation of several association rules with a confidence value above 0.3 indicates that there is a fairly strong relationship between products.

Google Laptop and Cell Phone Stickers products appear as the dominant consequence in some rules. This indicates that the product has a high tendency to be purchased along with various types of stickers and other technology-themed accessories. In business, these findings can be used for cross-selling strategies and product bundling. For example, a rule with a confidence of 0.471 indicates a considerable chance of recommending Google Laptop and Cell Phone Stickers to consumers who buy Android Sticker Sheet Ultra Removable. This strategy has the potential to increase average order value and promotion effectiveness.

In addition, the linkage between products in similar categories, such as stickers and digital accessories, indicates the existence of theme-based purchasing patterns or consumer interest in technology merchandise. This is in line with the concept of market basket analysis, where consumers tend to buy products that have similar functions or categories in one transaction [15]. From a technical point of view, the use of a minimum support of 0.005 succeeds in filtering out fairly representative product combinations without producing too many rules. Meanwhile, a minimum confidence of 0.05 allows for the identification of diverse association patterns, which can then be selected based on their strength level.

Overall, the results of this study prove that the application of the apriori algorithm to Google Merchandise Store transaction data is able to produce valuable information in supporting data-based marketing decision-making. The association rules obtained can be implemented in the system of product recommendations, promotion grouping, and product display arrangement on e-commerce platforms.

4. CONCLUSION

Based on the results of research that has been conducted on 12,267 transactions with 382 product attributes using the apriori algorithm, seven association rules were obtained that meet the minimum support value of 0.005 and the minimum confidence of 0.05. The results of the analysis show that there is a pattern of interrelated purchases between products on the Google Merchandise Store, especially in the category of technology-themed stickers and accessories. The rule with the highest confidence value is Android Sticker Sheet Ultra Removable → Google Laptop and Cell Phone Stickers at 0.471, which shows that 47.1% of consumers who bought the product also bought Google Laptop and Cell Phone Stickers in the same transaction.

These findings indicate that there are opportunities to implement cross-selling and product bundling strategies to increase transaction value. Apriori algorithms have been proven to be able to identify frequent itemsets and form relevant association rules on e-commerce datasets with high product variety. The results of this research can be used as a basis for the development of a product recommendation system and the preparation of data-based marketing strategies.

REFERENCES

- [1] H. H. M. Altarturi, A. R. M. Nor, N. I. Jaafar, and N. B. Anuar, "A bibliometric and content analysis of technological advancement applications in agricultural e-commerce," *Electron. Commer. Res.* 2023 252, vol. 25, no. 2, pp. 805–848, Jan. 2023, doi: 10.1007/s10660-023-09670-z.
- [2] H. M. Masfer, M. A. Helmi, <https://dialnet.unirioja.es/servlet/autor?codigo=6853711>, and <https://dialnet.unirioja.es/servlet/autor?codigo=6853712>, "The role of digital marketing in shaping consumer behavior in e-commerce platforms," *Int. J. Prof. Bus. Rev. Int. J. Prof. Bus. Rev.*, ISSN 2525-3654, ISSN-e 2525-3654, Vol. 10, N°. 3, 2025 (Ejemplar Dedic. a Contin. Publ. e05371), vol. 10, no. 3, p. 14, 2025, doi: 10.26668/businessreview/2025.v10i3.5420.
- [3] W. Shan, "Optimal pricing strategies and decision-making systems in e-commerce using integrated fuzzy multi-criteria method," *Expert Syst.*, vol. 42, no. 1, p. e13468, Jan. 2025, doi: 10.1111/exsy.13468.
- [4] E. Dritsas and M. Trigka, "Machine Learning in E-Commerce: Trends, Applications, and Future Challenges," *IEEE Access*, vol. 13, pp. 99048–99067, 2025, doi: 10.1109/ACCESS.2025.3572865.
- [5] B. Rolando and H. Mulyono, "E-Commerce as a Catalyst for Digital Economy Development: A Study of Marketing Strategies and Their Impact," *J. Distrib. Sci.*, vol. 23, no. 4, pp. 61–79, 2025, doi: 10.15722/jds.23.04.202504.61.
- [6] I. D. Hunyadi, N. Constantinescu, and O. A. Țicleanu, "Efficient Discovery of Association Rules in E-Commerce: Comparing Candidate Generation and Pattern Growth Techniques," *Appl. Sci.* 2025, Vol. 15, Page 5498, vol. 15, no. 10, p. 5498, May 2025, doi: 10.3390/app15105498.
- [7] S. Wu, "An association rule-based approach for frequent item mining of multi-stage access data," *Discov. Comput.* 2025 281, vol. 28, no. 1, pp. 139–, Jul. 2025, doi: 10.1007/s10791-025-09644-9.
- [8] P. Anusha, G. Sripriya, S. Samreen Mohammed, B. Abhinay, and N. Thrishool, "International Journal of Communication Networks and Information Security Leveraging Market Basket Analysis for Targeted Promotions and Personalized Recommendations," 2025, Accessed: Feb. 18, 2026. [Online]. Available: <https://https://ijcnis.org/>
- [9] H. A. Maulana and A. N. Rohman, "Apriori-Based Association Rule Mining Approach for Developing a Product Recommendation System in an Agricultural E-Marketplace," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 4, pp. 566–572, Oct. 2025, doi: 10.32736/sisfokom.v14i4.2486.
- [10] N. Mazumdar and P. K. D. Sarma, "Sequential pattern mining algorithms and their applications: a technical review," *Int. J. Data Sci. Anal.* 2024 203, vol. 20, no. 3, pp. 1683–1726, Oct. 2024, doi: 10.1007/s41060-024-00659-x.
- [11] H. Lin, W. Gan, G. Huang, and P. S. Yu, "Towards high-utility sequential rules with repetitive items," *Int. J. Mach. Learn. Cybern.* 2025 1610, vol. 16, no. 10, pp. 7729–7745, Jun. 2025, doi: 10.1007/s13042-025-02683-0.
- [12] A. L. Macready *et al.*, "Why trust is crucial – The moderating role of trust in the relationship between motivation and intention to buy healthy, sustainable and novel foods," *Food Qual. Prefer.*, vol. 126, no. 3, p. 105386, May 2025, doi: 10.1016/j.foodqual.2024.105386.
- [13] T. R. Siam, M. A. Mozumder, S. A. Sakib, M. S. Arefin, and A. W. Reza, "Mining Association Rules for Optimizing Customer Purchase Behavior in E-Commerce Transactions," 2025 *IEEE Int. Conf. Quantum Photonics, Artif. Intell. Networking, QPAIN 2025*, 2025, doi: 10.1109/QPAIN66474.2025.11172084.
- [14] P. H. Putra, D. Selvida, and M. S. Novelan, "Application of Apriori Algorithm in Data Mining to Find Consumer Purchasing Patterns in Supermarkets," *J. Komput. Teknol. Inf. Sist. Inf.*, vol. 4, no. 1, pp. 221–226, Jun. 2025, doi: 10.62712/juktisi.v4i1.392.
- [15] G. Damaschi, A. Aboueldahab, and M. D'Addario, "Decomposing Brand Loyalty: An Examination of Loyalty Subcomponents, Product Price Range, Consumer Personality, and Willingness to Pay," *Behav. Sci. (Basel)*, vol. 15, no. 2, p. 189, Feb. 2025, doi: 10.3390/bs15020189.