

Using the Apriori Algorithm to Identify Purchase Patterns for Enhancing Sales in Personal Shopper Services

Euis Fadilah^{1*}, Ahmad Faqih², Sandy Eka Permana³

^{1,2} Informatics Engineering, STMIK IKMI Cirebon

³ Informatics Management, STMIK IKMI Cirebon
euisfadilahh5@gmail.com^{1*}

Abstract

This research aims to explore the application of the Apriori algorithm in identifying purchasing patterns in the drop-off service industry in order to increase sales. Drop-off services often face challenges in designing effective marketing strategies due to limited understanding of customer purchasing behavior. In this study, the Apriori algorithm is applied to uncover recurring purchase patterns among customers, which are then used to develop more efficient marketing strategies. Customer transaction data is analyzed to find associations that reflect their purchasing preferences. The results show that the application of the Apriori algorithm successfully identifies patterns that can improve marketing strategies and, ultimately, increase sales. This research emphasizes the importance of applying data mining techniques to improve the performance of delivery services.

Keywords: Apriori Algorithm, Purchasing Patterns, Delivery Service, Data Mining, Marketing Strategy.

1. Introduction

The rapid development of information technology has had a major impact on various sectors, including business and public services. One of the innovations that has emerged in this digital era is delivery services, which facilitate the transaction process between consumers and service providers. This service is increasingly in demand because it offers convenience, speed, and comfort for its users. However, in the midst of increasingly fierce competition, the biggest challenge for service providers is how to optimize sales and services. One aspect that needs to be considered is the utilization of transaction data, which is increasingly abundant, but often poorly managed.

A major problem faced by many companies, including checkout providers, is how to manage and analyze data to find patterns that can help increase sales. As technology develops, data mining algorithms, such as *Apriori*, are becoming increasingly important in analyzing large and complex transaction data. However, while these algorithms are widely used in various sectors, their application in trust services is still limited. Many service providers have yet to fully utilize their data to formulate more effective business strategies, so the potential for increased sales is often not maximized. The development of technology can be seen from the various forms of retail that have used the system in carrying out the buying and selling process. For example, in making buying and selling transactions, data on purchased goods is stored in one database. So that if data processing is not done properly, it will cause data accumulation and uselessness [1]

This research aims to implement the Apriori algorithm in finding purchase patterns in the transaction data of deposit service. By using the Apriori algorithm, it is expected to find associations between products that are often purchased together, which can then be used to formulate strategies to increase sales. This approach aims to identify hidden patterns in transaction data that can help service providers in designing more efficient and data-driven marketing strategies. With the existence of data mining, knowledge will be obtained in a large collection of data according to [2] The application of Apriori Algorithm, in general, requires a data structure to store candidate frequent itemsets for a kth iteration and to store the resulting frequent itemsets by [3] Application of Apriori Algorithm in Data Mining to Predict Visitor Patterns at Karo Regency Tourism Objects to [4]

2. Research Methods

The process of completing this research involves several stages, which include the following:

1. Preparation for Research; The initial stage of this research includes determining the background, identifying problems, and determining the boundaries of problems related to deposit services, which aims to provide a clear direction in the further research process.
2. Literature Review; This method is carried out by the author by reading and or citing books, journals and other sources related to modeling using data mining techniques and the Apriori algorithm [5].
3. Research Problem; This stage aims to understand the main problems in the delivery service by identifying purchasing patterns. The information obtained will be used to formulate relevant solutions to optimize service performance.
4. KDD (Knowledge Discovery in Databases) is a structured process for extracting useful information for decision-making, starting from data selection to result evaluation:
 - a. Selection; In this stage, the relevant dataset is selected using an Excel file with 4,999 records of secondary data from a transaction database. The "Reactive" operator is used to ensure that the selected dataset aligns with the research objectives and contains the necessary information for further analysis. The data selection process is the selection of relevant data for research. In the initial database, there are many attributes in the table that are not needed because they are not used in the data mining process[6]
 - b. Preprocessing; The pre-processing stage is the initial step in dataset processing, where transaction data from the service delivery system is cleaned to ensure its quality. This process includes removing irrelevant data, correcting errors, and reducing noise. Attribute transformation, especially for Customer ID and Product, is also an important part of this stage.
 - c. Transformation; To process the dataset with the Apriori algorithm, the data must be transformed into a binomial format. This is done using the Numerical to Binomial operator to enable the algorithm to accurately identify association patterns. In this case, identifying the delivery service is also crucial, as the attributes related to the service need to be adjusted to match the binomial format. This stage is the process of transforming the selected data as a result of synchronizing the previous data for the data mining process[7].
 - d. Data Mining; The FP-Growth operator is used to analyze transaction data by identifying association patterns between items that are frequently purchased together. This algorithm optimizes the search for frequent itemsets using the FP (Frequent Pattern) tree structure, enabling efficient processing of large datasets. In the context of delivery services, FP-Growth helps identify items that frequently appear together in delivery transactions, which can be used for targeted promotional strategies or more accurate product recommendations.
 - e. Evaluation; The patterns discovered from the Apriori algorithm are analyzed to identify associations between products that are frequently purchased together. Evaluation is carried out by observing the support, confidence, and lift values to ensure the significance of the identified patterns. The results are translated into sales strategies, such as product bundle recommendations or promotions

2.1. Data Mining

Data Mining, The process of extracting and searching for useful information in selected data using specific techniques or methods depending on the knowledge-seeking process and the purpose of the research by [8]. Data mining is a term used to describe the discovery of knowledge in databases. The process involves analysis to discover unexpected relationships and summarize data in a way that is more understandable and useful to the data owner. In this study, the researchers used a quantitative approach by applying the Knowledge Discovery in Databases (KDD) process, which focuses on extracting knowledge from large datasets. This process was combined with the Apriori algorithm, which is known to be effective in identifying association patterns between items in transaction data. Using the KDD approach, researchers explored relevant purchase patterns, which can provide important insights into consumer behavior in delivery services. The results of applying this algorithm are expected to help service providers to optimize marketing strategies and improve service efficiency based on customer preferences and habits. The association data mining method can identify relationships or rules related to customer purchasing behavior and assist companies in making decisions regarding sales and marketing strategies[9] Association data mining methods can identify relationships or rules related to customer purchasing behavior buying behavior and help companies to make decisions regarding sales and marketing strategies[10]. Data mining is not an entirely new field. One of the difficulties in defining data mining is the fact that it inherits many aspects and techniques from established fields[11].

2.2. Apriori algorithm

The Apriori algorithm is an algorithm used to identify association patterns and relationships between items in a dataset. Its application can be used to find items that are often purchased together, which can then be used as the basis for promotional strategies, in the hope of increasing the purchase of additional items that were not previously planned. The a priori algorithm is a basic algorithm that was proposed by Agrawal & Srikant in 1994 to determine Frequent itemsets for Boolean association rules. Apriori algorithm is a type of association rule in data mining by[12]

One of the stages of association analysis that has attracted the attention of many researchers in developing efficient algorithms is frequent pattern mining. The significance of an association can be measured by two criteria: support and confidence. Support is the percentage of occurrences of an item combination in the database, while confidence is the strength of the relationship between items in the association rule[13]

The most commonly used data mining method is the Apriori technique by crosschecking the frequency of data sets using the apriori method. The Apriori method uses the concept of frequencies with symbols that are already understood beforehand to process the next information[14]. Association rule is a technique used to identify relationships or interesting patterns between elements in a dataset, as well as to understand how the presence of one item can influence the occurrence of another item. This technique is commonly used to discover associations between items in data. The Apriori algorithm, as one of the algorithms in association analysis, is used to find relationships between two or more attributes or objects in a dataset. In this context, the Apriori algorithm plays a key role in implementing association rules in data mining. Generally, the process of association analysis consists of two main stages, which are:

4	0	1	0	0	0	0	0	0	1	0	1	0	1	0	1
5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
8	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
9	0	0	1	0	1	0	1	0	1	0	1	0	0	0	0
10	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
12	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
14	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
15	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0
18	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0
19	0	1	1	1	0	0	0	0	0	0	0	0	1	0	1
20	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
...
4998	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0
Jumlah	1026	1038	970	1056	1041	1014	1023	988	983	1001	1001	947	1013	956	918

To identify purchase patterns in delivery services, the apriori calculation is used, which includes data selection and cleaning, as well as transformation into the appropriate format. The apriori algorithm searches for frequently bought item combinations based on support and calculates confidence to find relationships between items. The results can be used for product recommendations or promotional strategies.

To identify purchasing patterns in delivery services to increase sales, the Apriori calculation is used. This Apriori calculation involves several processes, namely:

1. Itemset Arrangement

This solution begins with the formation of 1 itemset, which consists of items that appear significantly in the transactions, using the concept of minimum support. In this study, the minimum support value is set at 30%, meaning only itemsets that appear in at least 30% of the transactions will be considered. Itemsets that meet this criterion will be used for further analysis, such as discovering association rules using the Apriori algorithm. Itemsets with support below 30% will be excluded from further processing.

Table 3: Result Itemset Arrangement

Name Items	Amount	Support (dalam persen)
Adidas Sneakers	1026	20.5
Barbie doll	1038	20.8
Faber Castell Sketchbook	970	19.4
Funko Pop Batman	1056	21.1
Harry Potter book set	1041	20.8
La Roche-Posay Cleanser	1014	20.3
Lego city set	1023	20.5
Levi's Jeans	988	19.8
Loreal Foundation	983	19.7
Maybelline Mascara	1001	20
Pilot G2 Pen Set	1001	20
The Body Shop Lotion	947	18.9
Uniqlo T-Shirt	1013	20.3
Wardah Lipstick	956	19.1
Zara dress	4082	81.6

2. Combination of Two Items

The process of forming C2, also known as the 2-itemset, is carried out with a minimum support of 30%. Table 4 below shows the 2-itemset combinations calculated using the following formula:

Combinations of 2 itemsets that do not meet the minimum support threshold will be excluded.

Table 4: Result Combination of Two Items

Name Items	Amount	Support (dalam persen)
Zara dress, Funko Pop Batman	880	17.6
Zara dress, Harry Potter Book Set	862	17.2
Zara dress, Barbie Doll	870	17.4
Zara dress, Adidas Sneakers	845	16.9
Zara dress, Lego city set	840	16.8
Zara dress, la roche-posay cleanser	830	16.6
Zara dress, Uniqlo T-Shirt	835	16.7

Zara dress, Maybelline Mascara	830	16.6
Zara dress, Pilot G2 Set	820	16.4
Zara dress, Levi's Jeans	810	16.3
Zara dress, Loreal Foundation	800	16.2
Zara dress, Faber Castell Sketchbook	795	15.9
Zara dress, Wardah Lipstick	810	16.3
Zara dress, The body shop lotion	775	15.5

3.2. Flowchart

In analyzing this research, the author did 2 (two) ways, namely the first way, manually by creating several itemset tables in finding association rules according to the rules of the apriori algorithm[15]. The design of the flowchart can be explained as follows:



Fig. 1: Flowchart

Process flowchart steps:

1. In the picture above, it is explained that it starts with start,
2. A preliminary study is carried out to obtain a basic understanding, explore the problem being studied and prepare a basis for further research.
3. Then proceed with data processing by applying Knowledge Discovery in Databases; selection, preprocessing, transformation, data mining, and evaluation to find patterns and information from large amounts of data.
5. Analyzing data using Rapid Miner software
6. Furthermore, the results of association rules are created
7. At the end completed

4. Testing

In the testing stage, the researchers used RapidMiner software to analyze the questionnaire data with a minimum support of 30% and minimum confidence of 50%. A 30% support means that only itemsets that appear in at least 30% of the transactions are considered, while a 50% confidence indicates that an association rule is only considered if the likelihood of purchasing the second item is 50% or higher after purchasing the first item.

The results showed that with a 30% minimum support, Zara Dress had the highest results in terms of association, meaning it was frequently purchased together with other products in the transactions. Therefore, this product was selected to proceed to the next step, which is the 2-item combination, to explore further relationships with other products. These findings provide valuable insights for designing more targeted promotional strategies, such as bundling products that are often bought together into one package or special promotion.

No.	Premises	Conclusion	Support	Confidence
3	la roche-posay cleanser	zara dress	0.166	0.819
4	faber castell sketchbook	zara dress	0.159	0.819
5	lego city set	zara dress	0.168	0.820
6	loreal foundation	zara dress	0.162	0.822
7	adidas sneakers	zara dress	0.169	0.823
8	uniqlo t-shirt	zara dress	0.167	0.825
9	harry potter book set	zara dress	0.172	0.828
10	maybelline mascara	zara dress	0.166	0.828
11	levi's jeans	zara dress	0.164	0.831
12	funko pop batman	zara dress	0.176	0.832
13	barbie doll	zara dress	0.174	0.836
14	wardah lipstick	zara dress	0.163	0.850

Fig. 2: Result data Association Rules

4.1. Formation of Association Rules

After successfully identifying all the high-frequency patterns in the dataset, the next step is to perform testing using RapidMiner software to find association rules based on these frequency patterns. This process aims to explore relationships or associations between items that frequently co-occur in transactions, which provides important insights into the purchasing patterns that occur in the dataset. By performing this analysis, we can gain a deeper understanding of how certain items relate to each other in transactions, as well as identify products or items that have a high chance of being purchased together. The results of this association rule mining process will be presented in the form of frequency patterns that describe the combination of related items, as well as the relationship between items based on how often they appear together in various transactions that occur. As a final result, 14 association rules were found that describe important patterns in purchasing, which can be used as a basis for designing more effective and efficient marketing strategies.

AssociationRules

```

Association Rules
[pilot g2 pen set] --> [zara dress] (confidence: 0.814)
[the body shop lotion] --> [zara dress] (confidence: 0.817)
[la roche-posay cleanser] --> [zara dress] (confidence: 0.819)
[faber castell sketchbook] --> [zara dress] (confidence: 0.819)
[lego city set] --> [zara dress] (confidence: 0.820)
[loreal foundation] --> [zara dress] (confidence: 0.822)
[adidas sneakers] --> [zara dress] (confidence: 0.823)
[uniqlo t-shirt] --> [zara dress] (confidence: 0.825)
[harry potter book set] --> [zara dress] (confidence: 0.828)
[maybelline mascara] --> [zara dress] (confidence: 0.828)
[levi's jeans] --> [zara dress] (confidence: 0.831)
[funko pop batman] --> [zara dress] (confidence: 0.832)
[barbie doll] --> [zara dress] (confidence: 0.836)
[wardah lipstick] --> [zara dress] (confidence: 0.850)
    
```

Fig. 3: Result Association Rules

Figure 3 is the result of testing using the rapid miner application application by using a support value of 30% and confidence of 50%. Seen in this figure has the same rules as the previous calculation

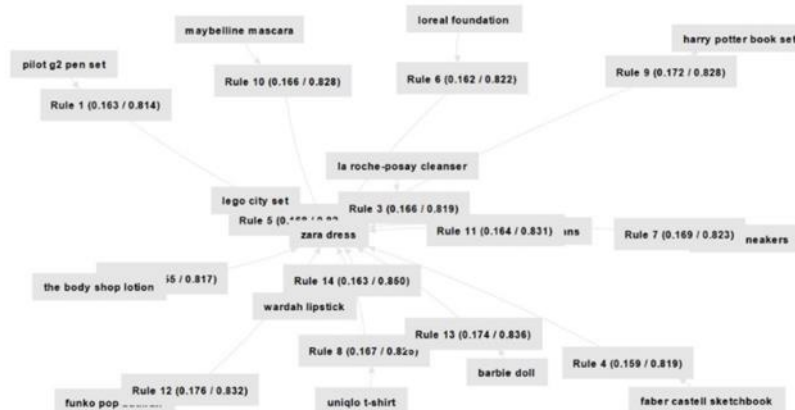


Fig. 4: Graph Association Rules

5. Conclusion

This research successfully identifies customer purchasing patterns in a drop-off service using the Apriori algorithm, which can increase sales through data-driven marketing strategies. Apriori algorithm is effective in finding product associations that are often bought together, which provides insights to design more targeted promotions.

The results of the analysis show that products such as Zara Dress have the opportunity to be purchased with other products, such as Pilot G2 Pet Set and The Body Shop Lotion. These findings support the development of promotional strategies, such as product bundling, cross-selling, and data-driven inventory management. With this approach, delivery services can improve customer experience, loyalty, and competitiveness in the market.

References

- [1] I. H. Sigalingging and R. Harman, "Analisis Pola Pembelian Produk Menggunakan Algoritma Apriori Iska," *Build. Informatics, Technol. Sci.*, vol. 2, no. 1, pp. 17–22, 2020, [Online]. Available: <https://ejurnal.seminar-id.com/index.php/bits/article/view/285>
- [2] N. Sari, D. B. Soedijino, and A. Nasiri, "Penerapan Metode Algoritma Apriori untuk Menentukan Pola Pembelian Produk," *Sisfotenika*, vol. 11, no. 1, pp. 1–13, 2020.
- [3] F. A. Sianturi, "Penerapan Algoritma Apriori Untuk Penentuan Tingkat Pesanan," *J. Mantik Penusa*, vol. 2, no. 1, pp. 50–57, 2018, [Online]. Available: <http://bowmasbow.blogspot.com/20>
- [4] S. Sinaga and A. M. Husein, "Penerapan Algoritma Apriori dalam Data Mining untuk Memprediksi Pola Pengunjung pada Objek Wisata Kabupaten Karo," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 49–54, 2019, doi: 10.34012/jutikomp.v2i1.461.
- [5] Romdani and S. Rahmatullah, "Analisis Pola Pembelian Konsumen Menggunakan Algoritma Apriori Pada Data Transaksi Penjualan Apotek & Alkes 58," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 5, pp. 790–798, 2022, doi: 10.32672/jnkti.v5i5.4995.
- [6] N. Hadinata and Kurniawan, "Analisis Pola Pembelian Produk Makanan Ringan Menggunakan Algoritma Apriori," *J. SISFOKOM (Sistem Inf. dan Komputer)*, vol. 9, no. 1, pp. 537–546, Mar. 2023, doi: 10.37012/jtik.v9i1.1614.
- [7] S. Amanda, D. Setiawan, and L. Trisnawati, "Penerapan Algoritma Apriori Dalam Menganalisis Pola Minat Beli Konsumen Di Coffee Shop," *JEKIN - J. Tek. Inform.*, vol. 3, no. 1, pp. 25–32, 2023, doi: 10.58794/jekin.v3i1.483.
- [8] S. Komariyah, Saeful Anwar, and Bani Nurhakim, "Implementasi Data Mining FP-Growth Untuk Analisis Pola Pembelian Pada Transaksi Penjualan," *J. Manaj. Dan Bisnis Ekon.*, vol. 1, no. 2, pp. 62–75, 2023, doi: 10.54066/jmbe-itb.v1i2.128.
- [9] R. Destriyanah, K. Kaslani, E. Wahyudin, G. Dwilestari, and M. Mulyawan, "Penerapan Algoritma Fp-Growth Untuk Menentukan Pola Pembelian Makanan Di Warmindo," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2159–2165, 2024, doi: 10.36040/jati.v8i2.8969.
- [10] A. Musaddad, O. Nurdiawan, and G. Dwilestari, "Penerapan Association Rule Menggunakan Frequent Pattern Growth Untuk Rekomendasi Produk Jersey Sepakbola," *J. Comput. Syst. Informatics*, vol. 3, no. 3, pp. 100–105, 2022, doi: 10.47065/josyc.v3i3.1390.
- [11] R. Rismanto, L. Darmawan, and A. Prasetyo, "Penerapan Algoritma Apriori Dalam Menentukan Pola Pembelian Konsumen Di Kafe Hidden Toast and Float," *J. Teknol. Inf. dan Terap.*, vol. 4, no. 2, pp. 83–88, 2019, doi: 10.25047/jtit.v4i2.64.
- [12] R. Yudha, S., K. Auliasari, and Prasetya, P., R., "Penjualan Produk Bangunan," *J. Mhs. Tek. Inform.*, vol. 4, no. 1, pp. 154–161, 2020.
- [13] Andika, M., A., N. Suarna, and R. D. Dana, "Analisa Dataset Asosiasi Penjualan Menggunakan Metode FP-Growth," *J. Teknol. Ilmu Komput.*, vol. 2, no. 1, pp. 80–88, 2023, doi: 10.56854/jtik.v2i1.108.
- [14] I. Zulfa, R. Rayuwati, and K. Koko, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode apriori," *Tek. J. Sains dan Teknol.*, vol. 16, no. 1, p. 69, 2020, doi: 10.36055/tjst.v16i1.7601.
- [15] S. Aliyah, "Analisis Pola Pembelian Toys Dengan Menggunakan Metode Algoritma Apriori," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 472–478, 2020.