Apriori Algorithm Analysis of Purchase Pattern Prediction on CV. Askara Media Literature

Yoga Fitriana, Muhammad Syahputra Novelan, Muhammad Iqbal

Abstract

This study aims to analyze consumer purchasing patterns at CV. Askara Sastra Media by utilizing data mining techniques with the A priori algorithm. The data used consists of sales transaction data that has undergone preprocessing to ensure data quality and consistency. The Apriori method is applied to discover frequent itemsets and generate association rules based on support and confidence values. The results show that the A priori algorithm is capable of identifying relationships between products that are frequently purchased together, with several association rules having high confidence and lift values. The resulting purchasing patterns provide valuable insights for the company in determining marketing strategies, such as product bundling and product placement. Therefore, the implementation of the A priori algorithm is proven to be effective in supporting data-driven decision-making and improving sales strategy efficiency at CV. Askara Literature Media.

Keywords: *Data Mining, Apriori algorithm, Purchasing pattern, Association rule, Sales*

Yoga Fitriana[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: yoga.fitriana.yf@gmail.com[1]

Muhammad Syahputra Novelan[2], Muhammad Iqbal[3]
[2,3]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: putranovelan@dosen.pancabudi.ac.id[2], wakbalpb@yahoo.co.id[3]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

The rapid development of information technology has encouraged various companies to utilize data as a strategic asset in decision-making. One form of data utilization is through data mining techniques, which are able to dig up hidden information from large data sets. In a business context, stored sales transaction data is often only used as an archive, even though there are important patterns that can be used to improve marketing and sales strategies [1][2].

CV. Askara Sastra Media as a company engaged in content according to the business field, for example printing, publishing, or product distribution, has purchase transaction data that continues to grow along with operational activities. However, the data processing has not been optimally utilized to find out consumer purchasing patterns. This causes the company does not have a strong analytical basis in determining strategies such as product structuring, bundling promotion, and stock management.

One of the methods in data mining that can be used to find patterns of relationships between items in transactions is the A priori algorithm. This algorithm works by identifying frequent itemsets and forming association rules that show the relationship between products based on the level of support and confidence. By applying an A priori algorithm, companies can find out which products are often purchased simultaneously by consumers [3].

The application of a priori algorithmic analysis to transaction data in CV. Askara Sastra Media is expected to be able to provide new insights into customer purchasing patterns. This information can be used to improve the effectiveness of marketing strategies, such as more optimal product placement, product package offerings, and more accurate data-driven decision-making.

Companies that focus on distributing or selling goods need an effective strategy in setting sales, promotion, and stock management packages. Without proper data analysis, companies risk inefficiency in stock management and loss of opportunities to sell [5]. Previous research indicates that the use of algorithms a priori can improve the efficiency of marketing strategies through product suggestions based on purchasing patterns. In addition, the incorporation of other methods such as K-Nearest Neighbor can be used to improve the accuracy of predicting consumer shopping patterns. However, the application of the method must be adjusted to the characteristics of the data and the needs of the company to produce appropriate and relevant information [6][7].

Based on these problems, this study aims to analyze the purchasing patterns of consumers in CV. Askara Sastra Media uses an A priori algorithm. Through this research, it is hoped that association patterns between products can be found that can be used as a basis for decision-making for business strategies, such as product structuring, bundling promotions, and improving customer service [8][9].

Literature Review

Data mining is the process of obtaining valuable information from large data sets by applying certain techniques to uncover hidden patterns. In the business world, data mining is often used to analyze consumer behavior, especially in finding purchasing patterns that can support strategic decisions[10]. One of the commonly used methods in data mining is association rule mining, which aims to identify the relationships between items in a transaction dataset. A priori algorithm is one of the well-known techniques in association rule mining that is used to identify frequent itemsets based on minimum support values and trusts. The basic principle of this algorithm is to take advantage of a priori properties, i.e. if an itemset does not meet the minimum support, then all supersets of that itemset will also not meet the criteria. In this way, the search process becomes more effective because it is able to reduce the number of itemet candidates that need to be examined[11][12].

Studies by Agrawal and Srikant show that a priori algorithm can generate efficient association rules in analyzing sales transaction data. In addition, other research shows that this algorithm can be applied to shopping cart analysis to identify products that are often purchased

at the same time by consumers[13]. This data is invaluable in marketing strategies such as product bundling arrangements and the layout of goods in stores. In addition to the A priori algorithm, a number of previous studies have also shown the success of applying the A priori algorithm in various sectors. A study by Hidayatullah shows that a priori use of sales data can result in association rules that favor improved marketing strategies. Another study by Nugroho also indicates that the application of a priori can produce appropriate product recommendations based on customer purchasing patterns [14][15]. In addition, a study by Rahmawati shows that the application of data mining in sales systems is able to improve data management efficiency and support data-based decision-making. Sari mentioned that association rule mining is very efficient in finding relationships between products in sales. From the literature research, it can be concluded that the A priori algorithm is an efficient way to analyze consumer spending patterns. By utilizing this algorithm, companies can access important information regarding the relationships between products that can be used to improve their business strategy [16]. Thus, this study uses an A priori algorithm to analyze purchasing patterns in CVs. Askara Sastra Media as a step to improve the effectiveness of data-based decision-making

Research Methodology

This study applies a quantitative approach through the data mining method to analyze consumer spending patterns in CV. Askara Literary Publications. The method used is an A priori algorithm, which is an association rule mining technique that aims to identify the relationship between items in sales transaction data. The data used is historical transaction data that includes information about the goods purchased by the consumer over a certain period of time. The research process begins with data collection, preprocessing, the formation of itemsets that often appear, to the creation of association rules which are then analyzed to support decisions in business. In the initial stage, sales transaction data is collected which is then followed by a preprocessing process to clean the data from duplication, incomplete data, and adjust the data format to be ready for analysis. The next data is converted into a binary table format to indicate the presence of an item in each transaction. After that, a frequent itemset is formed by calculating the support value for each item and combination of items. Support values are used to determine the frequency with which an item appears in all transactions, which is formulated as follows:

$$\textbf{\textit{Support(A)}} = \frac{Jumlah\ transaksi\ yang\ mengandung\ A}{Total\ transaksi} \quad (1)$$

In addition, to determine the appearance of a combination of two items, the calculation of the support of the pair of items is used which is formulated as follows:

$$\textbf{\textit{Support(A)}} = \cap\ \textbf{\textit{B}}\ \frac{Jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{Total\ transaksi} \quad (2)$$

Once the frequent itemset that meets the minimum support value is obtained, the next step is to establish an association rule by calculating the confidence value. The confidence value is used to measure the level of confidence that if an item A is purchased, then item B will also be purchased, which is formulated as follows:

$$\textbf{\textit{Confidence(A→B)}} = \frac{Support(A \cap B)}{Support(A)} \quad (3)$$

To strengthen the analysis, elevator ratio measurement is also used which aims to determine the strength of the relationship between items. The elevator value indicates whether the relationships between items are positive, independent, or negative, with the following formula:

$$\textbf{\textit{Lift(A→B)}} = \frac{Confidence(A \rightarrow B)}{Support(B)} \quad (4)$$

All the resulting association rules are then selected based on the minimum value of support and confidence that has been determined. Rules that meet these criteria will be analyzed to find significant consumer purchasing patterns. The results of this analysis are expected to provide strategic recommendations for companies in determining marketing strategies, product arrangements, and stock management of goods more effectively and efficiently. To clarify the

flow of research conducted, a research architecture is presented that describes the stages of the process from data processing to evaluation of results, as shown in Figure 1.

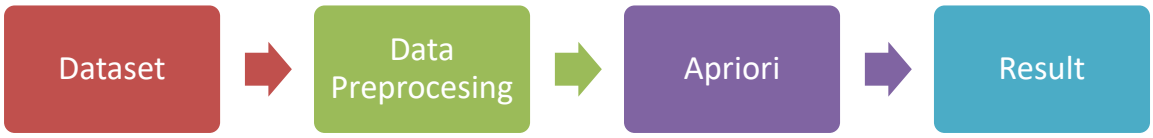

Figure 1. The architecture of the system of pattern analysis and purchase prediction is based on apriori.

Figure 1 shows the research architecture used in analyzing purchasing patterns and predicting purchasing categories on CVs. Askara Literature Media. This research consists of several main stages, namely data overview, data preprocessing, data transformation, application of a priori algorithm for purchase pattern analysis, application of K-Nearest Neighbor (KNN) algorithm for prediction, and evaluation and analysis of results.

Data Overview

This stage is the process of collecting and understanding initial data which includes sales transaction information such as Customer ID, book category, product price, and purchase date.

Data Preprocessing

At this stage, data cleaning is carried out, such as deleting duplicate data, handling missing values, and encoding and converting data types so that they are ready to be used in the analysis process.

Data Transformation

The transaction data is then transformed into a form suitable for the data mining process, namely by grouping transactions and converting the data into binary format using the one-hot encoding technique.

A priori (Purchase Pattern)

A priori algorithms are used to find buying patterns by generating frequent itemset and association rules based on support, confidence, and lift values.

Evaluation and Analysis

The final stage is an evaluation of the prediction results using a confusion matrix, accuracy, precision, recall, and F1-score, as well as an analysis of the resulting purchasing patterns.

Table 1 CV customer transaction dataset. Askara Literature Media.

	E-mail address	Purchase Date	Book Category
0	CUST00001	2025-03-11	[Technology]
1	CUST00001	2025-03-19	[Novel, Comics]
2	CUST00001	2025-05-23	[Novels, Educational Books, Comics]
3	CUST00002	2025-02-03	[Novel, Comics]
4	CUST00002	2025-03-24	[Novel, Comics]

The dataset used is CV customer transaction data. Askara Sastra Media which consists of Customer ID attributes, date of purchase, and book category. This data is used as a basis in the process of analyzing purchasing patterns.

Table 2 The result of transforming data into binary form (one-hot encoding).

	Educational Books	Comics	Motivation	Novels	History	Technology
0	False	False	False	False	False	True
1	False	True	False	True	False	False
2	True	True	False	True	False	False
3	False	True	False	True	False	False

4	False	True	False	True	False	False

The transaction data is then transformed into binary form using the one-hot encoding technique, where each category of the book is represented in True (bought) and False (not bought) values. This transformation is needed to simplify the analysis process using a priori algorithm.

Table 3 The frequency of occurrence of each category of books.

	0
Educational Books	3057
Comics	2209
Motivation	1448
Novels	3001
History	1461
Technology	1490

Next, the frequency of occurrence of each item is calculated to determine the initial appearance rate of each category of books in the dataset.

Table 4 The result of the formation of the Frequent Itemset is based on the support value.

	Support	**Item Sets**
0	0.440046	(Educational Books)
1	0.317979	(Comics)
2	0.208435	(Motivation)
3	0.431985	(Novel)
4	0.210307	(History)

Support values are calculated to determine how often an item appears in the dataset. Items that meet the minimum support value will be used in the formation of association rules.

Results

The findings of this study come from the analysis of sales transaction data on CVs. Askara Sastra Media utilizes an A priori algorithm. The data used is transaction data that has passed the preprocessing stage so that it is ready to be analyzed in binary table format. Based on this information, the process of forming a frequent itemset is carried out by setting the minimum value of support and confidence that has been determined in advance. In the initial phase, the support value for each item is calculated. Calculations show that some items have a high support value, signifying that they appear frequently in transactions. Items that have a support value

above the minimum limit are then selected as a frequent 1-itemset. Next, the combination between items is performed to form a 2-itemset and a 3-itemset, then the support value is recalculated to assess whether the combination meets the criteria as a frequent itemset. From the results of the analysis, a number of combinations of items with high support values were obtained. This shows the tendency of consumers to buy those items at the same time. For example, a combination of certain products shows a significant support value, so it can be classified as a frequent buying pattern. After the itemset is frequently obtained, the next step is to create an association rule by calculating the trust value. The results of the calculation show that some rules have a high level of confidence, which indicates a strong relationship between items. The association rules with the highest trust values indicate that if a consumer buys a certain product, then it is more likely to buy another item. In addition, elevator ratio measurements are also carried out to understand the strength of the relationship between items. The analysis showed that the majority of the resulting rules had an elevator value above 1, which signified a positive relationship between the items. This shows that the combination of products has an important relationship and can be used in marketing strategies

In general, the findings of this study suggest that a priori algorithm can effectively recognize consumer purchasing patterns. The resulting patterns can be used as a basis for business decision-making, such as the selection of promotional packages, product arrangements, and stock management. The application of the A priori algorithm, thus, makes an important contribution in increasing the efficiency and effectiveness of sales strategies in CVs. Askara Media Literature

Table 5. The resulting association rules use a priori algorithm.

	Background	Consequences	Support	Confidence	Elevator
0	Motivation	Educational Books	0.136894	0.656768	1.492498
1	Technology	Educational Books	0.139341	0.649664	1.476356
2	Comics	Novels	0.243846	0.766863	1.775207
3	Novels	Comics	0.243846	0.564479	1.775207
4	History	Novels	0.136174	0.647502	1.498898

Based on the results of processing using a priori algorithm, several association rules were obtained. The rule with the highest confidence score is *Comics → Novels* of 0.766863, which indicates that customers who buy comics are more likely to also buy novels. An elevator value of 1.775207 indicates a strong relationship between the two items.

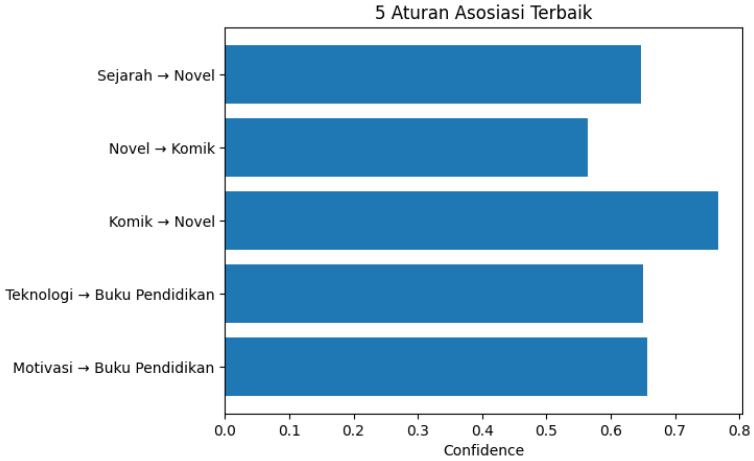

Figure 2. Visualization of the 5 Best Association Rules Based on Confidence Values

The graph shows the five best association rules based on confidence values. It can be seen that *the rules of the → Novel Comics* have the highest value compared to other rules.

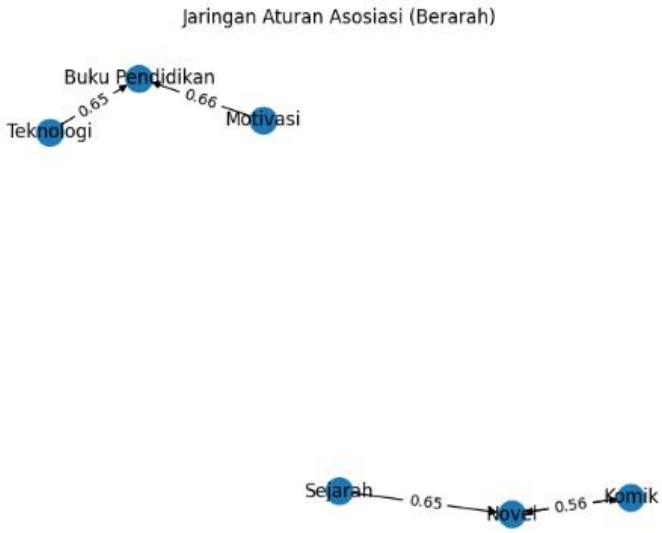

Figure 3. Visualization of the Network of Association Rules Between Book Categories

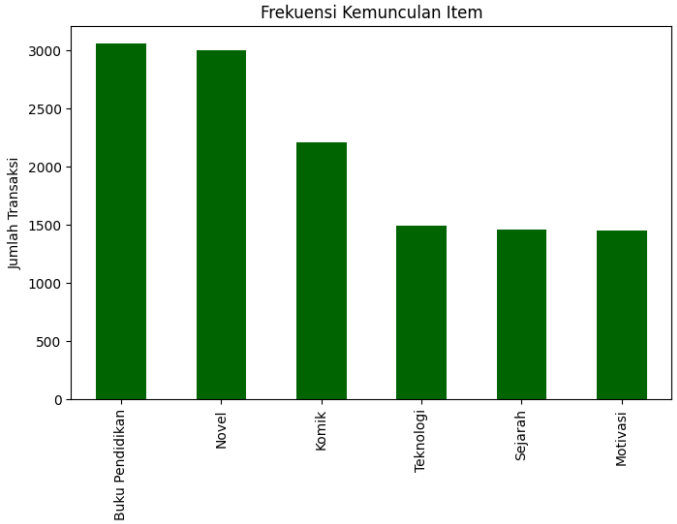

Figure 4. Visualization of the Network of Association Rules Between Book Categories

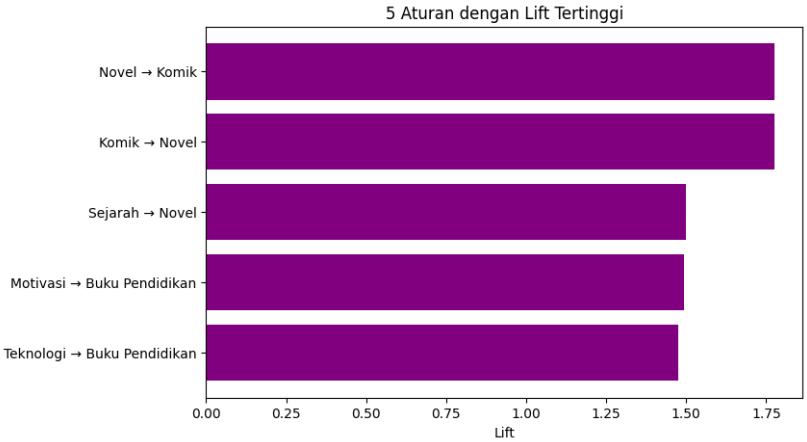

Figure 5. Visualization of the Network of Association Rules Between Book Categories

Based on the results of visualization of a priori algorithm analysis on CV. In Askara Sastra Media, the first graph shows the frequency of item appearances or how often each category of books is purchased in a transaction. It can be seen that Educational Books have the highest number of transactions, followed by Novels and Comics, which indicates that these three

categories are the main products that customers are most interested in. Meanwhile, categories such as Technology, History, and Motivation have a lower frequency, but remain significant as a complement in purchasing patterns.

The second chart shows the 5 association rules with the highest lift value, illustrating the strength of the relationships between items in a single transaction. Rules like Comic Novels and Comic Novels have the highest lift value around 1.78, which means these two categories are very often bought together and have a strong relationship. Additionally, patterns such as Novel History and Educational Book Motivation and Technology show that customers who purchase certain books are likely to follow the purchase pattern of continuing to other categories. Overall, these results show that there is a tendency for category-based purchases, so they can be leveraged for business strategies such as product bundling, automated recommendations, and catalog structuring to increase sales.

Conclusion

From the results of the research that has been conducted, it is concluded that the use of a priori algorithm can identify consumer spending patterns in CVs. Ask Askara Sastra Media in an effective way. By analyzing transaction data, frequent itemsets and association rules are obtained that indicate the relationship of products that are often purchased at the same time. The support and confidence values obtained show that the A priori algorithm can be used to identify important and relevant purchasing patterns in supporting the business decision-making process. In addition, evaluation using the lift ratio showed that most association rules had values above one, which signified a positive relationship between items. This confirms that the patterns found indicate a high level of relationships and can be used for marketing strategies, such as product bundling determination, product structuring, and increased promotional effectiveness.

The implications of this study show that companies can leverage purchasing pattern analysis to improve operational efficiency and data-driven sales strategies. By understanding consumer spending trends, companies can better manage stock and improve customer satisfaction through more accurate product recommendations. For further research development, it is recommended that the A priori algorithm be combined with other methods such as K-Nearest Neighbor to improve accuracy in predicting purchasing patterns. In addition, the use of a wider and more diverse dataset can result in a more in-depth and precise analysis.

References

[1] A. H. Anshor, A. A. Sulaeman, dan S. Winarni, “Association Rule to Increase Sales Using the Apriori Algorithm Method,” vol. 4, no. 1, hal. 321–331, 2024.

[2] M. Sadikin, D. Ridha, D. Putri, M. Reza, dan R. O. Batubara, “Implementasi Metode Association Rule Mining dalam Mencari Pola Penjualan Sim Card Selular Menggunakan Algoritma Apriori,” vol. 1, no. 1, 2021.

[3] D. Dwiputra, A. M. Widodo, H. Akbar, G. Firmansyah, dan U. E. Unggul, “EVALUATING THE PERFORMANCE OF ASSOCIATION RULES IN APRIORI AND FP-GROWTH ALGORITHMS : MARKET BASKET ANALYSIS TO DISCOVERRULES OF ITEM COMBINATIONS,” vol. 2, no. 8, hal. 1229–1248, 2023, doi: 10.58344/jws.v2i8.403.

[5] F. S. Zikri *et al.*, “THE COMPARISON BETWEEN THE APRIORI ALGORITHM AND THE FP-GROWTH ALGORITHM IN DETERMINING FREQUENT PATTERN PERBANDINGAN ANTARA ALGORITMA APRIORI DENGAN ALGORITMA FP-GROWTH DALAM MENENTUKAN,” vol. 10, no. 2, hal. 615–625, 2025.

[6] M. Al-maolegi dan B. Arkok, “A N I MPROVED A PRIORI A LGORITHM FOR,” vol. 3, no. 1, hal. 21–29, 2014.

[7] M. E. Rana, “Apriori Algorithm based Association Rule Mining to Enhance Small - Scale Retailer Sales,” *2023 IEEE 6th Int. Conf. Big Data Artif. Intell.*, hal. 187–191, 2023, doi: 10.1109/BDAI59165.2023.10256952.

[8] R. A. Putra, M. Amalia, M. Putri, dan S. M. Sinaga, "Implementation of Association Rules Algorithm to Identify Popular Topping Combinations in Orders," vol. 1, no. January, hal. 95–101, 2024.
[9] T. Pajak, P. Di, H. Rodhiy, dan Z. Sitorus, "Data Mining Menggunakan Algoritma Apriori Dalam Menentukan," vol. 4, no. 2, hal. 198–204, 2023.
[10] P. Studi, S. Informasi, P. S. Informatika, P. Studi, dan S. Informasi, "Analisis Dan Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Penjualan Pada Kantin Universitas Tanri Abeng," vol. 12, no. 3, hal. 210–218, 2019, doi: 10.30998/faktorexacta.v12i3.4541.
[11] M. Bhargava dan A. Selwal, "Available Online at www.ijarcs.info Association Rule mining using Apriori Algorithm : A Review," vol. 4, no. 2, hal. 2–5, 2013.
[12] D. Prayogi, M. S. Novelan, S. R. Lubis, M. A. Rizko, dan A. Guna, "Analisis Pola Pembelian Konsumen Menggunakan Algoritma Apriori dan," vol. 4, no. 2, hal. 285–290, 2025.
[13] M. H. Santoso, "Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom," vol. 1, no. 2, hal. 54–66, 2021.
[14] D. M. Tank, "Improved Apriori Algorithm for Mining Association Rules," no. June, hal. 15–23, 2014, doi: 10.5815/ijitcs.2014.07.03.
[15] H. B. Sabila dan A. A. Algorithm, "Implementation of Apriori Algorithm for Data Mining on Sales Transaction Data," vol. 6, no. 3, hal. 189–193, 2023.
[16] N. S. Poli dan A. S. Sikder, "Predictive Analysis of Sales Using the Apriori Algorithm : A Comprehensive Study on Sales Forecasting and Business Strategies in the Retail Industry," no. 1, hal. 1–15, 2023.