Journal of Applied Informatics and Computing (JAIC)
Vol. 10, No. 1, February 2026, pp. 566–574
e-ISSN: 2548-6861

Comparison of Apriori and FP-Growth Algorithms in Market Basket Analysis for Online Book Sales

Sofia Rizkal Karimah [1] *, Erika Devi Udayanti [2] *
* Information Systems, Dian Nuswantoro University
112202206893@mhs.dinus.ac.id [1], erikadevi@dsn.dinus.ac.id [2]

Article Info

Article history:

Received 2025-12-02
Revised 2025-12-30
Accepted 2026-01-07

Keywords:

Association Rule Mining,
Apriori,
FP-Growth,
Market Basket Analysis,
Book Sales.

ABSTRACT

This research aims to compare the performance of the Apriori and *FP-Growth* algorithms in the process of data mining association patterns in the online sales transaction data of a bookstore. The dataset used consists of 74.090 transactions resulting from data cleaning from the period January-June 2025. The analysis was conducted through the stages of data collection, followed by data preparation consisting of data cleaning and data transformation, and then continued to the modeling stage of the two algorithms. The results of the experiment show that Apriori tends to be faster on small-scale datasets with simple transaction patterns, while FP-Growth has more stable memory usage and shows more efficient processing time on parameters that analyze larger data. Both algorithms produce identical numbers and contents of association rules for each parameter variation, indicating that the significant difference lies in performance efficiency, and not in the knowledge patterns produced. Rules with the highest *lift* values, such as the association between the books "Rumah Kaca" and "Jejak Langkah" (lift: 183,306 & *confidence* 0,903) and between the books "Namaku Alam" and "Pulang" (lift: 34,062 & *confidence*: 0,51) indicate strong purchasing patterns between titles with the same author and theme. These findings have the potential to support *cross-selling* strategies and product recommendations in online sales systems. This research is still limited to a relatively small and homogeneous dataset, so further using a broader data coverage is recommended to test the algorithm's performance more comprehensively.

I. Introduction

Advances in information technology and the era of digitalization have changed the pattern of interaction between consumers and businesses, including in the book sales industry. In an era of increasingly popular online shopping, the large number of transactions that occur every day generates big data that has the potential to provide strategic information for companies if analyzed correctly [1]. This requires bookstores to participate in the development of business strategies that are suitable for buyer behavior. Currently, recommendation systems for online book sales are not yet optimal, with buyers/users tending to be recommended general books, without knowing the patterns formed in buyer characteristics. Therefore, an analytical approach is needed that can explore the information contained in existing online sales transaction data.

One analytical approach that is widely used to explore hidden patterns in transaction data is Market Basket Analysis (MBA). Through MBA, companies can identify the relationship between products that are often purchased together, thereby supporting the implementation of *cross-selling* strategies and product recommendations based on customer behavior [2] [3]. MBA is a relevant technique for improving the quality of recommendations, especially when combined with algorithms that can accelerate the computation process on big data [4].

MBA techniques are rooted in the *concept of Association Rule Mining* (ARM), which aims to find relationships between items in the form of *if–then* rules based on support, confidence, and lift values. This approach enables more objective decision making because it is based on statistically proven patterns from actual transaction data [5] [6]. One of the functions of association rules mining is its ability to extract frequent itemsets that describe customer purchasing habits, which can then be interpreted as association rules.

The two most commonly used algorithms in ARM are Apriori and Frequent Pattern Growth (FP-Growth). The Apriori algorithm uses a candidate generation approach to gradually form item combinations, while FP-Growth builds an FP-Tree data structure that compresses transaction data to find patterns without explicit candidate generation [7]. In general, FP-Growth is known to be faster in handling large datasets because it only requires two database scans, but the effectiveness of both can vary depending on the dataset size, support threshold value, and the complexity of the pattern being searched for [8] [9].

Recent studies indicate that the performance difference between the two algorithms is not always significant. In research [2], which evaluated the Apriori, Apriori TID, and FP-Growth algorithms by dividing the dataset into four seasons, FP-Growth was found to be superior in terms of time efficiency for datasets with a high number of transactions, while Apriori remained relevant and easy to interpret for small-scale datasets.

Similar results were reported by [10], with producing 22 rules with 100% confidence and 1% support, stating that both algorithms can produce identical association rules, even though the minimum support and confidence parameter experiments were set with many variants. The main differences were only in the duration of the execution process and the amount of memory used by Apriori.

The results of the [6] research, which took a dataset from a large pharmaceutical company with a data range of 6 months of sales, also confirmed that FP-Growth tends to be superior with large data and only takes half the time of the Apriori algorithm. However, when applied to small data, Apriori is more stable and transparent with simple data with low pattern complexity. Considering that both algorithms produce the same rules.

Thus, most previous studies have focused on time and memory efficiency. However, not many studies have applied this comparison to the domain of online book sales, where in today's global era, online sales are very prevalent and are a source of sales data that is worth researched. Also, transaction characteristics in this sector tends to have different levels of product variety and purchase frequency compared to general/groceries retail. In addition, the book sales sector can also be studied. With its high diversity of titles and rapidly changing literacy trends, the book sales domain is a relevant context to explore in order to identify purchasing patterns that can support recommendation strategies and sales development. .

Therefore, this research was conducted to analyze and compare the performance of the Apriori and FP-Growth algorithms in finding customer purchasing patterns in online bookstore sales data, with a focus on processing time efficiency and the number of rules formed in the association rules results. The results of this research are expected to provide empirical contributions to the application of association analysis in the context of the book retail business and serve as a reference for data-driven decision making.

II. METHOD

This research aims to analyze customer characteristics in the book sales business by utilizing online sales transaction data obtained from a bookstore retailer, to be ed using Google Colab, with the aim of implementing a cross-selling strategy in sales. The was conducted by comparing two algorithms that form the basis of Market Basket Analysis, namely the Apriori algorithm and the FP-Growth algorithm, to identify patterns of product purchases that often occur simultaneously, so that the results can support marketing decision making [2] [11]. The stages of the are shown in Figure 1 as a guide to the analysis process.

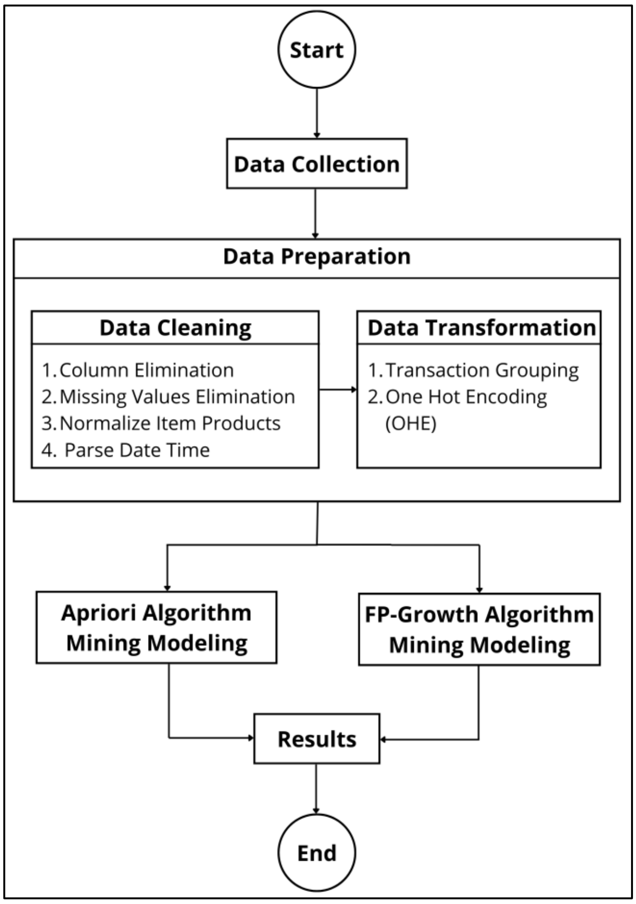

Figure 1. Method Flow

A. Data Collection

Online sales data for January-June 2025 was collected, with 49 columns and 295.454 transactions. Of the 49 columns of data used, 3 columns were the main focus of the research, namely NoTrans, ItemDesc, and PaymentTime.

Exploratory Data Analysis (EDA) was conducted to evaluate sales patterns and understand the characteristics of the data used [12]. It was revealed that at least 294.454 transactions occurred. The distribution of transactions is shown in Figure 2.

The result shown in Figure 2 showed that the highest number of transactions occurred on Monday, with 52.026 transactions. Followed by transactions on Friday with a total of 46.197 transactions, then on Sunday with 41.407 transactions, Tuesday with 40.183 transactions, Wednesday with a total of 39,140 transactions, Thursday with 38.112 transactions, and Saturday was the day with the lowest number of transactions with 37.389 transactions.

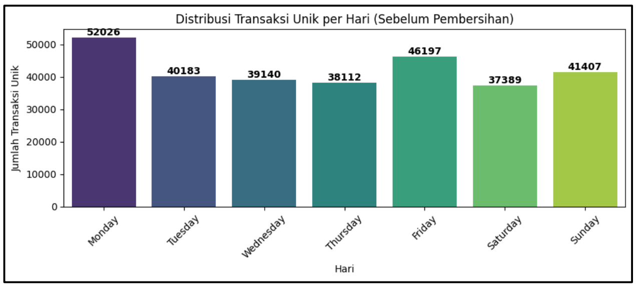

Figure 2. Bar Chart of the Daily Distribution of Transactions Using EDA

Based with this, the additional information is revealed that the average number of transactions occurring per day is 42.065 transactions.

B. Data Preparation

Data preparation was carried out in several stages, namely data cleaning and data transformation. This step ensures that the data is clean, uniform, and not dominated by noise, which would reduce the optimization of algorithm performance and avoid bias [13], and is ready to be used for association rule analysis and data exploration.

C. Data Cleaning

The data pre-processing stage began with cleaning the raw data obtained from the transaction files. Because the dataset was too large, unused columns were eliminated. Of the 294.454 transactions that occurred, it was found that the data was dominated by transactions that contained only one product. Although the dataset did not have missing values, a step was still taken to eliminate missing values in rows.

Transactions consisting of only one product were deleted in the market basket analysis because they could not form a minimal combination itemset (≥2 items), which is the basis for rule formation. The results of this process showed a reduction in sales data and data structure. The sales data, which was originally 294.454 transactions, became 74.090 transactions. A summary of all the changes that occurred can be seen in Table I.

TABLE I
COMPARISON OF DATASETS BEFORE & AFTER DATA CLEANING

Note	Number of Transactions	Number of Rows	Number of Columns	Total of Missing Values
Before Cleaning	294.454	450.376	49	0
After Cleaning	74.090	245.032	3	0

This process is followed by standardizing the text format in the product description column (ItemDesc) and converting the time data type to datetime format.

After the data cleaning stage was carried out, further analysis reveals that the cleaned dataset consists of 74.090 valid transactions and 14.482 unique book items. On average, each transaction contains approximately 2,43 items, indicating a relatively sparse transactional structure with small basket sizes. It should be noted that this 2,43 figure represents a statistical mean rather than a literal purchase quantity, which may explain the sparse transactional structure.

In terms of temporal distribution, transactions are unevenly distributed across days, with higher transaction volumes occurring on weekdays compared to weekends, and it can be seen in Figure 1. The combination of high item cardinality and low average basket size increases the overall complexity of frequent itemset generation, as a large number of possible item combinations must be evaluated despite limited co-occurrence within individual transactions. These characteristics highlight the computational challenges inherent in mining association rules from real-world online retail transaction data.

D. Data Transformation

The data structure in the dataset is still in row data breakdown. Each row in the NoTrans column that has the same value means that in that 1 NoTrans, the customer purchased more than 1 different product. An example of this data can be seen in Table II.

Table II
DATASET BEFORE GROUPING BY TRANSACTION

NoTrans	ItemDesc	…	PaymentTime
1	Kemari Kota 02	…	01/01/2025 00:05
1	Attack on Titan Bind Up Edition 04	…	01/01/2025 00:05
…	…	…	
74090	Atomic Habits	…	30/06/2025 23:58
74090	The Psychology of Money	…	30/06/2025 23:58

The transaction grouping stage is carried out to convert the cleaned data into a transaction basket format [14]. Each NoTrans is represented as a single transaction containing a list of products (Product Name) purchased together. This process also removes duplicate products in the same order to meet the requirements of the association rules algorithm. This grouping resulted in 74.090 valid transactions, and examples of transactions containing the Order ID (NoTrans), product list (ItemDesc), and purchase date (PaymentTime) can be seen in Table III.

Table III
DATASET AFTER PER TRANSACTION GROUPING

NoTrans	ItemDesc	…	PaymentTime
1	['Kemari Kota 02', 'Attack on Titan Bind Up Edition 04', 'Naruto Bind Up Edition 06']	…	1/1/2025 00:05
2	['Funiculi Funicula', 'Sword Art Online 005 Phantom Bullet', 'Sword Art Online 006 Phantom Bullet', 'Hai, Miiko! 35', 'Hai, Miiko! 36']	…	1/1/2025 00:06
…	…	…	…
74090	['Atomic Habits', 'The Psychology of Money', 'The E Myth Revisited']	…	30/6/2025 23:58

After the grouping process, the dataset represents multi-item purchase behavior with sufficient transactional density. On average, each transaction contains more than two items, and average of 2indicating adequate interaction among products to support meaningful association rule mining analysis.

In the next stage, this research applied the One-Hot Encoding (OHE) technique to convert transaction data, which was originally in the form of a list of products in each transaction, into a binary matrix format. Each column represents a specific product, while each row is a unique transaction marked with a value of 1 if the product was purchased and 0 if it was not involved in the transaction. This binary representation is required because frequent itemset mining algorithms such as Apriori and FP-Growth operate on transactional data in binary form. The one-hot encoding process ensures that all transactions are represented uniformly, enabling accurate identification of frequent itemsets and association rules [15]. The OHE approach also improves the accuracy of frequent itemset formation and minimizes information loss in retail transaction data [4][16].

Table IV
RESULTS OF ONE HOT ENCODING FOR SOLD PRODUCTS

No Trans	Kemari Kota 02	Attack on Titan Bind Up Edition 04	…	Hai Miiko! 35	Hai Miiko! 36
1	1	1	…	0	0
2	0	0	...	1	1
…	…	…	…	..	…
74090	0	0	…	0	0

E. Association Rules Mining Modeling

Entering the core stage of this research, we find patterns of relationships between products that are often purchased together by customers. In this modeling, we will compare two algorithms, the Apriori algorithm and the FP-Growth algorithm.

All experiments were conducted using the Google Colab platform in a CPU-based execution environment with Python as the programming language. Data preprocessing, transaction grouping, and exploratory data analysis were performed using the pandas and NumPy libraries. The Apriori and FP-Growth algorithms were implemented using the frequent pattern mining functions provided by the mlxtend library. Execution time measurements were used to support the comparative evaluation of computational efficiency under identical experimental settings.

This stage begins with finding frequent itemsets that occur in the selected dataset using the two algorithms. After the frequent itemsets are found, modeling is performed to evaluate the differences in the performance of the two algorithms and determine which of the two algorithms produces the best results.

The algorithm in ARM generates patterns with if-then results, which means that if a customer buys product A, then it is highly likely that they will buy product B [17].

In the formation of association rules, there are two things that become benchmarks, namely:

1) *Support*: or the supporting value is the percentage of item combinations sold in the dataset [18]. The *support* value is crucial because it shows how strong the item is in *the* overall transaction dataset.

To find the support value of an item, the following equation is used:

$$\text{Support (A)} = \frac{\text{Total number transaction of item A}}{\text{Total number of transactions}} \quad (1)$$

To find the support value of two items, the following equation is used:

$$Support\ (A, B) = \frac{\text{Total number transaction of item A \& B}}{\text{Total number of transactions}} \quad (1)$$

2) *Confidence*: or certainty value. This value explains how often buyers who purchase product A are also likely to purchase product B. The confidence value explains how strong the relationship between items is in the association rule [18].

$$Confidence\ (A, B) = \frac{\text{Total transactions of item A \& B}}{\text{Total Transaction Consist Item A}} \quad (2)$$

3) *Lift Ratio*: as a benchmark value for the relationship between products A and B compared to the relationship between random products. After the support and confidence values are found, the validity of the relationship between a combination of items A and B must be tested [19]. An association rule is considered valid if its lift ratio value is greater than 1. To calculate the lift ratio, the Support B value

is needed as the consequent (consequence/follower). The Support B value is calculated in the same way using equation (1).

The lift ratio value is then determined using the equation below:

$$Lift\ Ratio\ (A,B) = \frac{Confidence\ (A,B)}{Support\ B} \qquad (3)$$

In frequently occurring itemsets, the support results may be the same, but the confidence values may differ, and the combinations may be very different. To find strong associative rules, it is necessary to find associative rules where the dependence level of the antecedent (trigger) and consequent items meets the minimum confidence requirement. In this research, several stages will be carried out to determine the optimal values for support and confidence.

The first step is to find the optimal support from the top 10 frequent itemsets, by trying each support value. The second step is to perform several trial-and- s to determine the support that produces the most relevant association rules. The fourth step is to determine the best confidence value. The final step is to evaluate the performance of the Apriori and FP-growth algorithms that can provide maximum association rule results for business decisions and deliver the best performance for continuous use with company datasets.

In this study, multiple minimum support values were evaluated to identify meaningful association patterns. Based on the experimental results, a minimum support of 0.004 and a minimum confidence of 0.4 were selected as optimal parameters, balancing rule relevance and computational efficiency for both Apriori and FP-Growth algorithms

III. Results and Discussion

A. Results

In this section, the results of the dataset modeling that is ready for analysis using two algorithms will be explained.

1) *Association Rules Mining Modeling:* In the data, several items were found to be most frequently purchased by buyers. The top 10 items purchased can be seen in Table V.

TABLE V
ITEMSETS RESULTS USING APRIORI AND FP-GROWTH

itemsets_str	support
Seporsi Mie Ayam Sebelum Mati	0,043920
The Psychology of Money	0,035295
Atomic Habits	0,026225
Laut Bercerita	0,025779
Sisi Tergelap Surga	0,023013
Filosofi Teras	0,020664
Bicara Itu Ada Seninya	0,019409
Seorang Pria yang Melalui Duka dengan Mencuci Piring	0,017749
Sebuah Seni untuk Bersikap Bodo Amat	0,015805
Sang Alkemis	0,015711

After determining the support values of the 10 products most frequently purchased by customers, as shown in Table V, the second step is to find the optimal support values that can form the association rules.

Table V shows that the optimal support range is between 0,04 and 0,01. However, in the 0,04-0,009 minimum support experiment, no rules were generated, and only a difference in computation time was found, with the Apriori algorithm being faster than FP-Growth in that minimum support range, as shown in Table VI. Therefore, the parameter was finally lowered to further testing. These differences are quantitatively evaluated based on processing time and memory usage metrics, as reported in Tables VI and VII.

Tables VI and VII show a clear performance trend between the two algorithms. When the minimum support value is relatively high and no association rules are generated, Apriori demonstrates slightly faster execution times due to its simpler candidate generation process on shallow data. However, as the minimum support decreases and the number of generated rules increases, Apriori experiences a significant rise in memory usage and processing time.

In contrast, FP-Growth maintains stable memory consumption and exhibits faster execution times when mining denser frequent itemsets. These results indicate that FP-Growth scales more efficiently under lower support thresholds, while Apriori remains suitable for small-scale or shallow transaction datasets.

TABLE VI
COMPARISON OF PROCESSING TIME BETWEEN APRIORI AND FP-GROWTH

Min, Support	Rules	Apriori		FP-Growth	
		Memory (MB)	Process time	Memory (MB)	Process time
0,009	0	49,76	0,962	1,21	1,394
0,01	0	37,04	1,281	1,22	1,623
0,02	0	3,97	1,179	1,21	1,904
0,03	0	1,89	1,216	1,21	2,230
0,04	0	1,81	0,981	1,21	1,742

Based on the experimental results in Table VII, it was determined that the support value to be used is 0,004. At this value, the rules do not produce patterns that are too rare or overfitted, and can still reveal the patterns of product associations that are most frequently purchased by customers.

TABLE VII
EXPERIMENTAL RESULTS WITH MIN. CONFIDENCE 0.1 AND MIN. SUPPORT VARIATIONS USING APRIORI & FP-GROWTH ALGORITHM THAT SHOWS SIGNIFICANCE CHANGES

Min. Support	Apriori			FP-Growth		
	Rules	Memory (MB)	Processing Time	Rules	Memory (MB)	Processing Time
0,008	0	69,69	0,995	0	1,34	0,832
0,007	2	105,94	1,145	2	1,47	1,007
0,006	6	183,32	2,205	6	1,67	0,879
0,005	8	292,92	2,2674	8	1,89	0,896
0,004	10	469,5	4,935	10	2,19	0,872
0,003	24	830,94	6,266	24	2,60	1,451
0,002	48	2596,81	18,615	48	3,64	2,056

Based on the results of Association Rules Mining modeling with a predetermined combination of support and confidence, the most balanced rule falls at minconf 0,4. This parameter has a fairly varied pattern and can reveal business value for implementation without having too low a confidence value.

TABLE VIII
EXPERIMENTAL RESULTS WITH MIN. *SUPPORT* 0.01 AND MIN. *CONFIDENCE* VARIANTS USING APRIORI & FP-GROWTH ALGORITHM

Min, Confidence	Number of Rules Formed	*Avg, Confidence*	*Avg, Lift*
0,3	5	0,6379	91,9317
0,4	5	0,6379	91,9317
0,5	3	0,7716	133,558
0,6	2	0,8981	183,306
0,7	2	0,8981	183,306
0,8	2	0,8981	183,306

Based on Tables IX and X, the results of the Apriori and FP-Growth algorithm analyses are exactly the same. With a minimum support parameter of 0,004 and a minimum confidence of 0,4, a number of association rules were obtained that show a strong relationship between book titles by the same author. The highest correlation pattern is seen in the rule of purchasing the books "Rumah Kaca" and "Jejak Langkah" with a confidence value of 0.903047. Buyers who purchased "Rumah Kaca" also purchased "Jejak Langkah", and had a support value of 0,4400 and a lift value of 183,306, which means that this pair of antecedents and consequents appears 183 times more often together than when the product is purchased with other products. This value indicates that the two books very often appear together in purchase transactions.

TABLE IX
ASSOCIATION RULES MINING RESULTS USING THE APRIORI ALGORITHM WITH MINSUPP 0.004 & MINCONF 0.4

Antecedents	Consequents	Support	Confidence	Lift
Rumah Kaca	Jejak Langkah	0,004400	0,903047	183,306
Jejak Langkah	Rumah Kaca	0,004400	0,893151	183,306
Namaku Alam	Pulang	0,006587	0,518597	34,062
Seorang Wanita yang Ingin Menjadi Pohon Semangka di Kehidupan Selanjutnya	Seorang Pria yang Melalui Duka dengan Mencuci Piring	0,005277	0,442308	24,920
Pulang	Namaku Alam	0,006587	0,432624	34,062

TABLE X
FP-GROWTH ALGORITHM *ASSOCIATION RULES MINING* RESULTS *MINSUPP* 0.004 & *MINCONF* 0.4

Antecedents	Consequents	Support	Confidence	Lift
Rumah Kaca	Jejak Langkah	0,004400	0,903047	183,306
Jejak Langkah	Rumah Kaca	0,004400	0,893151	183,306
Namaku Alam	Pulang	0,006587	0,518597	34,062
Seorang Wanita yang Ingin Menjadi Pohon Semangka di Kehidupan Selanjutnya	Seorang Pria yang Melalui Duka dengan Mencuci Piring	0,005277	0,442308	24,920
Pulang	Namaku Alam	0,006587	0,432624	34,062

Conversely, customers who purchased "Jejak Langkah" and also purchased "Rumah Kaca" experienced a decrease in confidence value to 0,893151. Although the support value remained the same and the lift value was still at 183,306, it was still categorized as very high.

On the other hand, the relationship between other books also shows a high association strength, such as in the rule for purchasing the books "Namaku Alam" and "Pulang" with a *support* value of 0,0066 , *confidence* of 0,5186 , and *lift* of 34,062. However, when the *antecedents* and consequents are reversed, the *confidence* value for this book rule decreases to 0,432624. With a lift value of 34.062 , it means that these two products appear together 34 times more often than when associated with other products.

Finally, the rule of the book "Seorang Wanita yang Ingin Menjadi Pohon Semangka di Kehidupan Selanjutnya" with the book "Seorang Pria yang Melalui Duka dengan Mencuci Piring" has a support of 0,0053 , *confidence* of 0,4423 , and *lift* of 24,920. All titles that appear in these parameters are from the same author and indicate a correlation in customer purchasing patterns.

B. Discussion

This section presents a comparative analysis of the performance of the Apriori and FP-Growth algorithms in the process of mining association patterns in online bookstore sales transaction data from January to June 2025. The discussion focuses on three main indicators, namely the efficiency of algorithm performance, the number of association rules formed in each parameter variation, and the level of similarity of the association rules generated by both algorithms. This analysis aims to assess not only the technical performance of both algorithms but also the consistency of the knowledge generated, thereby avoiding potential bias in performance assessment that focuses solely on processing time.

1) *Algorithm Performance Efficiency*: Based on the frequent itemsets experiment results in Table V, when the dataset was analyzed with optimal minimum support, the algorithm implementation on the dataset did not show significant results, as can be seen in Table VI. In fact, this parameter resulted in slow FP-Growth performance efficiency, even though memory usage was very stable. These results indicate that the parameters are too strict and shallow for the dataset, so that the FP-Growth algorithm tends to run longer than Apriori. On the other hand, the stable memory usage of FP-Growth is due to the data compression mechanism using the FP-Tree structure, which keeps the transaction representation size constant even if the support threshold value changes.

In Apriori, when run on a simple and small dataset, it can work a fraction of a second faster than FP-Growth. The increase in memory usage when Apriori is run occurs because the candidate generation process is performed repeatedly for each level of item combination, which results in greater memory consumption, especially when the support threshold decreases.

However, when the min. support is reduced and several rules are found that can be analyzed, as can be seen in Table VII, the FP-Growth computation time shows a significant change. The execution time of FP-Growth tends to be 50% faster than Apriori, while the stability of memory usage remains the same as in the experiment in Table VI.

These findings are supported by studies [3] and [20], which state that the FP-Growth algorithm works very well on large and variable datasets, while the Apriori algorithm operates optimally when the dataset is small and shallow.

2) *Formed Association Rules*: Based on the results of testing the min support and min conf parameters, it was found that both algorithms produced an identical number of association rules for each parameter variation used (Table VI and Table VII). The number of rules increased as the support value decreased, where the lower the frequency threshold, the more item combinations qualified as frequent itemsets. This pattern shows that parameter variables have a greater influence on model output than the type of algorithm used. The similarity in the number of rules for each parameter value confirms that both Apriori and FP-Growth have equivalent pattern discovery capabilities when applied to the same dataset and configuration.

These results are consistent with studies [3] and [7], which show that the differences in output rules between the two algorithms are more dominant in terms of performance efficiency rather than in the variation or accuracy of the rules produced. [9] also confirms that when the support and confidence thresholds are set identically, both algorithms produce the same number of rules, with the main difference being only in execution time. Meanwhile, research [7] adds that the stability of the identical number of rules in both algorithms reflects the semantic consistency of the patterns found, so that the selection of algorithms can be more focused on process efficiency without affecting the final results of the knowledge obtained.

From a business perspective, the minimum support value of 0,004 and minimum confidence of 0,4 were chosen because the experimental results revealed five rules that were sufficiently varied and provided business insights that could be used as sales strategies. These parameters are balanced, not too high so that the data is too small to analyze. Also, the parameters are not too small so that the data analyzed is too broad. If the parameters are too small, the product associations that occur are not strong enough to be used as a basis for strategic decisions by management.

3) *Similarity of Business Rule Results*: The test results show that the Apriori and FP-Growth algorithms produce identical association rules for each combination of minimum support and confidence values used. This similarity can be seen in the main metric values, namely support, confidence, and lift, which are exactly the same in each rule formed. This condition proves that even though the two algorithms have different processing mechanisms Apriori with a candidate generation approach and FP-Growth with a pattern-tree construction approach both produce equivalent knowledge output when the parameters and input data used are identical.

As presented in Table IX, the association patterns found have a very high level of relationship strength. The rule for the books "Rumah Kaca" & "Jejak Langkah" has a support of 0.0044, confidence of 0,903, and lift of 183,306, indicating that customers who buy the book "Rumah Kaca" almost always buy "Jejak Langkah" at the same time.

The reverse rule, the books "Jejak Langkah" & "Rumah Kaca", shows a slight difference confidence value of 0,893 but the same lift value, indicating a two-way relationship between titles from the same author. The decreased confidence value shows that among the transactions that purchased "Jejak Langkah", 0,893 included the product "Rumah Kaca". The consistent lift value between these two products shows that the association pair of these two products is 183 times greater than that of other products.

It should be noted that an extremely high lift value, such as 183,306, may occur when two items have relatively low individual frequencies but consistently co-occur within the same transactions. In such cases, lift reflects the strength of relative association rather than absolute purchase probability. Previous researches such in research [21] and [22] have shown that lift is highly sensitive to marginal item probabilities and may become disproportionately large when items have low support but strong conditional dependence, which may introduce bias if interpreted without considering support and confidence values. Therefore, lift should be interpreted cautiously, particularly in homogeneous datasets where in this case, thematic or author-based similarities dominate purchasing behavior.

In this research, the association rule between book “Rumah Kaca” and “Jejak Langkah” yields a lift value of 183,306, accompanied by a support value of 0,0044 and confidence values exceeding 0,9. Statistically, this indicates that the rule is derived from a relatively small subset of transactions, approximately 0.44% of all transactions which is 325 transactions, rather than representing a frequent purchasing pattern across the entire dataset. Nevertheless, the high confidence values demonstrate strong conditional reliability, indicating that when one title is purchased, the other is highly likely to be purchased within the same transaction. Thus, the observed lift represents a strong relative association within a niche purchasing segment, rather than an indicator of high-volume sales dominance.

A similar relationship was also found in the book rules "Namaku Alam" & "Pulang" (support 0,0066; confidence 0,519; lift 34,063) and the books "Pulang" & "Namaku Alam" (same results for support and lift, withconfidence 0,433), which show a thematic connection between titles with similar narratives, even though the antecedents and consequents rules for "Pulang" & "Namaku Alam" have lower confidence values than when the rules are reversed. In addition, the rule for the books "Seorang Wanita yang Ingin Menjadi Pohon Semangka di Kehidupan Selanjutnya" & "Seorang Pria yang Melalui Duka dengan Mencuci Piring" (support 0.0053; confidence 0.442; lift 24.921) confirms the existence of a pattern of simultaneous purchases of literary works with similar writing styles.

The similarity of association rule results between the two algorithms indicates that differences in mining methods do not affect the knowledge patterns produced. This reinforces the conclusions of [9] and [3], which state that the rule outputs generated by Apriori and FP-Growth will be identical if the support and confidence thresholds are set consistently.

From a business perspective, the association rule results have significant strategic implications. The strong relationship between book titles from the results of this research, after further in the dataset, are books written by the same author and on the same theme. This shows the potential for implementing a cross-selling strategy, where customers who buy one title can be recommended other books that have a strong thematic relationship (theme and author). In addition, this pattern of interconnection can be used to design promotional packages (product bundling) or a "recommended similar books" feature in online catalog systems to place books by the same author side by side.

These findings are in line with [18] which states that the application of association rules in e-commerce recommendation systems can increase the average basket value of customers without significantly increasing algorithmic costs. Thus, the similarity of the results between the two algorithms not only indicates the consistency of the analytical model, but also reinforces the validity of its application in the context of data-driven decision making for online book retail businesses.

Based on the experimental results, several high-confidence association rules were identified, reflecting strong conditional relationships in customer purchasing behavior. In an online bookstore scenario, such rules can be directly implemented in a recommendation or cross-selling module, where relevant books are suggested dynamically during the purchasing process. These recommendations may be delivered in real time. For example, when a customer selects or purchases *“Rumah Kaca”*, the system can immediately recommend *“Jejak Langkah”* based on a confidence value of 0,903. Such real-time recommendations are expected to increase click-through rate (CTR) by presenting highly relevant items that align with the customer’s current interest. Moreover, by exposing customers to strongly associated books during checkout, the likelihood of additional purchases may increase, thereby improving average basket value.

This context-aware recommendation approach enhances the relevance of suggested items and conceptually supports higher user engagement and the potential for increased purchases of related books Although this study does not conduct a direct business simulation, the high-confidence association rules derived from actual transaction data provide a strong empirical basis for implementing recommendation and cross-selling strategies in online book retail systems.

IV. CONCLUSION

This research compares the Apriori and FP-Growth algorithms in mining association patterns in online bookstore transaction data. The analysis results show that FP-Growth has more stable memory usage and is excellent for analyzing large data to identify buyer habits. Meanwhile, Apriori tends to be faster on small datasets with shallow characteristics, such as transactions that only occur two to three times. Both algorithms produce identical numbers and contents of association rules for each parameter variation, indicating that the main difference lies in performance efficiency, not in the patterns produced.

Rules with the highest lift and confidence values, such as "Rumah Kaca & Jejak Langkah" (lift 183,306 & confidence 0.903) and when the antecedents and consequents of the rule are reversed (lift 183.306 & confidence 0,89), then the rule "Namaku Alam & Pulang" (lift 34,063) and the rule "A Woman Who Wants to Be a Watermelon Tree in Her Next Life and a Man Who Copes with Grief by Washing Dishes" (lift 24.920), show a strong purchasing correlation between titles from the same author and with the same theme. This product pattern can be utilized for cross-selling strategies and product recommendations in online sales systems to maximize recommendations for books by the same author and with the same theme when buyers make transactions. Thus, both Apriori and FP-Growth are equally effective, and their selection can be adjusted to the size and complexity of the dataset being analyzed, so that this can be implemented in a customer transaction-based recommendation system.

This research still has limitations in terms of the size and characteristics of the dataset used, where the transaction data analyzed is relatively homogeneous and does not fully represent more complex sales conditions. Such homogeneity may strengthen association patterns among books with similar authors or themes, while limiting the generalizability of the extracted rules to more diverse retail environments.

In the future, this research can be developed by using a larger dataset covering multiple time periods and more diverse book categories. Further studies may also explore the integration of other algorithms, such as Eclat or sequential pattern mining, to capture temporal purchasing behavior and sequential buying patterns. Performance testing with algorithms other than Apriori and FP-Growth in denser data scenarios can also provide a more comprehensive picture in future . In addition, the results *of association rules* can be integrated directly into the recommendation system prototype to test their impact on increasing sales and user behavior in real life.

ACKNOWLEDGMENTS

Authors would like to express our deepest gratitude to our colleagues and Dian Nuswantoro University for their assistance in ensuring the smooth running of this research paper.

REFERENCES

[1] Caroline, Yuswardi, and Y. U. Rofi'i, "Analysis of E-Commerce Purchase Patterns Using Big Data : An Integrative Approach to Understanding Consumer Behavior," *Int. J. Softw. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 352–364, 2023, doi: 10.35870/ijsecs.v3i3.1840.

[2] A. I. Idris *et al.*, "Comparison of Apriori , Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores," *IJICS (International J. Informatics Comput. Sci.*, vol. 6, no. 2, pp. 107–112, 2022, doi: 10.30865/ijics.v6i2.4535.

[3] D. S. Nugroho, N. Islahudin, V. Normasari, and S. Z. Al Hakiim, "Application of Market Basket Analysis (MBA) Data Mining using the Apriori and FP-Growth Association Methods for Bundling Strategies at Wan Caffeine Addict Yogyakarta," *JISI J. Integr. Sist. Ind.*, vol. 11, no. 1, pp. 121–134, 2024, doi: 10.24853/jisi.11.1.121-134.

[4] A. F. Haqqi, W. Purnomo, and A. N. Rusydi, "Market Basket Analysis menggunakan Algoritma FP-Growth untuk Strategi Penjualan (Studi Kasus : PT XYZ)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 7, pp. 1–14, 2025, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15107

[5] R. A. Machfudiyanto, J.-H. Chen, Y. Latief, T. S. N. Rachmawati, A. M. Arifai, and N. Firmansyah, "Applying Association Rule Mining to Explore Unsafe Behaviors in the Indonesian Construction Industry," *Sustain. J.*, vol. 15, no. 6, 20023.

[6] D. Dwiputra, A. M. Widodo, H. Akbar, G. Firmansyah, and U. E. Unggul, "Evaluating the Performance of Association Rules in Apriori and FP-Growth Algorithms: Market Basket Analysis to Discover Rules of Item Combinations," *JWS J. World Sci.*, vol. 2, no. 8, pp. 1229–1248, 2023, doi: 10.58344/jws.v2i8.403.

[7] E. W. Pujiharto, Kusrini, and A. Nasiri, "Comparative Analysis of The Performance of The Apriori, FP-Growth and Eclat Algorithms In Finding Frequency Patterns In The Ina-Cbg's Dataset," *Cogito Smart J.*, vol. 9, no. 2, pp. 340–354, 2023, doi: 10.31154/cogito.v9i2.547.340-354.

[8] M. Rosadi and M. S. Hasibuan, "Comparison Of Apriori And Fp-Growth Algorithms In Analysing Consumer Behavior Towards The Purchase Of Electronic Data," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 12, no. 2, pp. 399–408, 2024.

[9] K. I. Lestari, S. Putra, and E. Rilvani, "The Efficiency Problem of the Apriori Algorithm and FP-Growth Evaluation Based on Literature Studies," *J. Media Akad.*, vol. 3, no. 7, pp. 1–15, 2025, doi: 10.62281.

[10] M. Raihan and Sutisna, "Comparative Analysis of the Apriori and FP-Growth Algorithms to Determine Sales Strategies at Maestro Jakarta Cafe & Space," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 3147–3157, 2024, doi: 10.35870/jimik.v513.994.

[11] U. M. Wulandari, A. T. Suseno, and M. Kholilurrahman, "Market Basket Analysis Using FP-Growth and Apriori on Distro Store Sales Transaction," *MATICS J. Ilmu Komput. dan Teknol. Inf.*, vol. 17, no. 1, pp. 12–18, 2025, doi: 10.18860/mat.v17i1.28820.

[12] R. Regulagadda *et al.*, "Customer Personality Analysis using Segmentation and Exploratory Data Analysis," *IJISAE Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 4s, pp. 794–800, 2024, [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/3864

[13] A. F. Achmad, A. Rahim, and N. A. Verdikha, "Implementasi Data Mining Algoritma Appriori pada Data Transaksi PIK Store," *JIP (Jurnal Inform. Polinema)*, vol. 11, no. 2, pp. 203–212, 2023, [Online]. Available: https://pdfs.semanticscholar.org/a3a6/719520863011cd35a933b303ebcb41b4e04f.pdf

[14] A. Rahmadsyah and N. Mayasari, "Grouping Goods With The Association Rule Method Using A Priori Algorithms In Modern Retail Stores," *Int. J. Sci. Technol. Manag.*, vol. 3, no. 6, pp. 1527–1532, 2022, doi: 10.46729/ijstm.v3i6.637.

[15] I. D. Hunyadi, N. Constantinescu, and O.-A. Ticleanu, "Efficient Discovery of Association Rules in E-Commerce : Comparing Candidate Generation and Pattern Growth Techniques," *Appl. Sci.*, vol. 15, no. 5498, pp. 1–20, 2025, doi: 10.3390/app15105498.

[16] L. Ahmedi, E. Rexhepi, and E. Bytyçi, "Using Association Rule Mining to Enrich User Profiles with Research Paper Recommendation," *Int. J. Comput. Digit. Syst.*, vol. 11, no. 1, pp. 1141–1146, 2022, doi: 10.12785/ijcds/110192.

[17] A. H. Wibowo, M. Andivas, D. Harits, H. Windyatri, and D. S. Nugraha, "A Pattern Analysis on Goods Purchase Relationship for Minimarket ' s Customers by Using Association Rule - Market Basket Analysis Method (AR-MBA)," *J. Inf. Technol. Ampera*, vol. 4, no. 3, pp. 195–206, 2023, [Online]. Available: https://journal-computing.org/index.php/journal-ita/index

[18] S. A. Miranda, Fahrullah, and D. Kurniawan, "Implementasi Association Rule Dalam Menganalisis Data Penjualan Sheshop dengan Menggunakan Algoritma Apriori," *Metik J.*, vol. 6, no. 1, pp. 30–36, 2022, doi: 10.47002/metik.v6i1.342.

[19] M. C. Sugianto, A. M. S, and I. A. Musdar, "Implementasi Data Mining Menggunakan Algoritma Apriori pada Toko Rumah Skincare 88," *JTRISTE*, vol. 11, no. 1, pp. 25–38, 2024.

[20] R. Wahyuningsih, A. Suharsono, and N. Iriawan, "Comparison of Market Basket Analysis Method Using Apriori Algorithm, Frequent Pattern Growth (FP-Growth) and Equivalence Class Transformation (ECLAT) (Case Study: Supermarket 'X' Transaction Data for 2021)," *Bus. Financ. J.*, vol. 8, no. 2, pp. 192–201, 2021, doi: 10.33086/bfj.v8i2.5226.

[21] F. Petry and R. Yager, "Data Mining Using Association Rules for Intuitionistic Fuzzy Data," *Information*, vol. 14, no. 7, p. 372, 2023, doi: https://doi.org/10.3390/info14070372.

[22] J. M. Luna, M. Ondra, H. M. Fardoun, and S. Ventura, "Optimization of Quality Measures in Association Rule Mining : An Empirical Study," *Int. J. Comput. Intell. Syst.*, vol. 12, no. 1, pp. 59–78, 2018, doi: 10.2991/ijcis.2018.25905182.