Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

June 2026. Vol. 5. No. 3; e-ISSN: 2808-4519

Implementation of the Apriori Algorithm for User Access Pattern Analysis on Debian Web Server

Jahanra Girsang[1]*, Lolo Ate tumangger [2], Lotar Mateus sinaga[3]

[1,2,3] *Teknik Informatika, Universitas Katolik Santo Thomas, Jalan Setia Budi No. 479 F, Tanjung Sari, Kecamatan Medan Selayang, Kota Medan, Sumatera Utara.*

jahanragirsang92@gmail.com[1], tumanggerlolo05@gmail.coml[2], lotarmateus88@gmail.com[3]*

Abstract

The increasing number of website user activities generates large amounts of log data that are often underutilized as a source of information for analyzing user access patterns. This study aims to implement the Apriori algorit hm to analyze user access patterns based on Apache Access Log data collected from a website running on Debian 12 using Apache Web Server and WordPress**.** The research data were obtained from the *access.log* file and processed through several preprocessing stages, including filtering, cleaning, and transforming the log data into transaction datasets based on user sessions. The Apriori algorithm was then applied to generate frequent itemsets and association rules using predefined minimum support and confidence values. The results show that the Apriori algorithm successfully identifies relationships among web pages that are frequently accessed together by users. The discovered patterns provide valuable insights into user navigation behavior, enabling website administrators to optimize website structure, improve service quality, and support data-driven decision-making. Therefore, the implementation of the Apriori algorithm on Apache Access Log data can serve as an effective approach for analyzing user behavior based on web server log data.

Keywords: Apache Access Log, Apriori Algorithm, Data Mining, Debian 12, Web Usage Mining, WordPress

1. Introduction

The rapid development of information technology has significantly increased the use of web-based services in various fields, including education, business, government, and public services. Websites have become the primary medium for delivering information, resulting in a continuous increase in user access activities. Every user activity is recorded by the web server in an access log file containing information such as IP addresses, access time, visited pages, request methods, and server response status. Although access logs contain valuable information, they are generally used only for system monitoring and maintenance purposes. As a result, their potential to provide insights into user behavior has not been fully utilized. Furthermore, the continuously growing volume of log data makes manual analysis inefficient and time-consuming [1].

One method that can be used to analyze log data is the data mining technique using the Apriori algorithm. The Apriori algorithm is an association rule mining method used to discover relationships or associations among items in a set of transactional data. In this study, each user browsing session on the website is treated as a transaction consisting of several accessed web pages. By applying the Apriori algorithm, frequent itemsets and association rules can be generated to describe user access patterns among website pages. This information can be utilized to understand user behavior, improve website navigation, optimize website structure, and support decision-making in web-based service management [2].

This study implements the Apriori algorithm on Apache Web Server access log data running on the Debian 12 operating system. The log data first undergoes a preprocessing stage to transform the raw data into transaction data suitable for analysis. The study focuses on identifying user access patterns based on relationships among web pages that are frequently visited within the same browsing session. The objectives of this study are to implement the Apriori algorithm for analyzing user access patterns on a Debian Web Server, identify relationships among web pages based on access log data, and analyze the utilization of association rule results to support website management and optimization. The results of this study are expected to serve as a reference for the application of data mining techniques in web server log analysis and to assist administrators in understanding user behavior more effectively [3].

2. Main body

2.1. Data Mining

Data mining is the process of extracting useful information from large datasets to discover patterns, relationships, trends, and new knowledge that were previously unknown. It integrates various disciplines, including statistics, artificial intelligence, machine learning, and database systems, to generate information that can support decision-making processes. The rapid development of information technology has led to a continuous increase in the amount of data stored in various systems. Such large volumes of data are difficult to analyze manually, making it necessary to employ methods capable of processing data automatically and efficiently. Data mining provides a solution by identifying hidden patterns within large datasets, thereby generating valuable information. In general, the data mining process consists of several stages, including data collection, data cleaning, data transformation, data analysis using specific algorithms, and interpretation of the obtained results. These stages are intended to ensure that the data used are of high quality so that the analysis results are reliable. One of the most widely used techniques in data mining is association rule mining, which is a method for discovering relationships among items within a dataset. This technique can identify items that frequently occur together and generate association rules that can be used to understand specific patterns in the data. In this study, the concept of data mining is applied to access log data generated by the Apache Web Server running on the Debian operating system. The log data, which records user activities, is processed into transactional data and subsequently analyzed using the Apriori algorithm. The results of this analysis are used to identify user access patterns to website pages, thereby supporting more effective website management and development [4].

2.2. Web Usage Mining

Web Usage Mining is a branch of data mining that focuses on analyzing website user activity data to discover behavioral patterns, browsing habits, and relationships among visited web pages. The data used in Web Usage Mining generally originate from web server log files, user clickstream data, browsing history, and other activity records generated while users interact with a website. The primary objective of Web Usage Mining is to gain insights into user behavior during website access so that the information can be used to improve service quality, enhance website navigation, and optimize the user experience. By identifying frequently occurring access patterns, website administrators can better understand user needs and make more informed decisions regarding website development. The Web Usage Mining process generally consists of several stages, namely data collection, data preprocessing, pattern discovery, and pattern analysis. During the data collection stage, information is obtained from log files generated by the web server. Next, preprocessing is performed to remove irrelevant data and transform the raw log data into a format suitable for analysis. Afterward, the pattern discovery stage is carried out using data mining techniques, such as the Apriori algorithm, to generate meaningful information. In this study, the primary data source is the *access.log* file generated by the Apache Web Server running on the Debian operating system. The log data are processed into a collection of transactions based on user browsing sessions and then analyzed using the Apriori algorithm to identify relationships among website pages that are frequently accessed together. The results of this analysis are expected to provide insights into user access patterns and support the development of a more effective website that better meets user needs [5].

2.3. Algoritma Apriori

The Apriori algorithm is one of the most widely used methods in association rule mining for discovering relationships or associations among items within a dataset. It was introduced as a technique for identifying frequent itemsets, which are groups of items that frequently occur together in a set of transactions. The resulting frequent itemsets are then used to generate association rules that describe the relationships among items. In this study, the Apriori algorithm is applied to analyze user access patterns on a website running on a Debian Web Server. Each user browsing session is treated as a transaction, while the web pages visited during the session are considered as items. Through this process, it is possible to identify web pages that are frequently accessed together by users. The fundamental principle of the Apriori algorithm is that if an itemset satisfies the minimum support threshold, then all of its subsets must also satisfy the minimum support threshold. Conversely, if an itemset does not satisfy the minimum support threshold, none of its larger supersets need to be analyzed. This principle enables the pattern discovery process to be performed more efficiently [6].

2.3.1. Frequent Itemset

Frequent itemsets are groups of items whose occurrence frequency exceeds a predetermined minimum support threshold. In this study, frequent itemsets are used to identify website pages that are frequently accessed together within a single user browsing session [7].

2.3.2. Support

Support is a measure used to determine the frequency of occurrence of an item or a combination of items within the entire transaction dataset.
The support formula can be expressed as follows:

$$\text{Support}(X) = \frac{\text{Number of transactions containing } X}{\text{Total number of transactions}} \quad (1)$$

Where:

- **X** represents an item or a combination of items being 5011nalysed.
- Support indicates how frequently the item or itemset appears in the entire transaction dataset.
- **A** higher support value indicates that the item or itemset occurs more frequently in the dataset.

For example, if the Home page appears in 80 out of 100 user browsing sessions, the support value of the page is calculated as follows:

$$\text{Support(Home)} = \frac{80}{100} = 0.8 = 80\% \quad (2)$$

2.3.3. Confidence

Confidence is used to measure the reliability of an association rule. It indicates the probability that a user will access page Y after accessing page X. The confidence formula is as follows:

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \quad (3)$$

Where:

- **X → Y** represents an association rule.
- **X ∪ Y** denotes the number of transactions containing both X and Y simultaneously.
- The confidence value indicates the strength of the relationship between item X and item Y.

As an example, if out of 50 users who accessed the Home page, 40 users also accessed the Product page, then:

$$\text{Confidence(Home} \rightarrow \text{Product)} = \frac{40}{50} = 0.8 = 80\% \quad (4)$$

The result indicates that 80% of users who accessed the Home page also accessed the Product page.

2.3.4. Association Rule

Association rules are the final output of the Apriori algorithm and are used to identify relationships among items within a transaction. An association rule is expressed in the form X → Y, which indicates that if item X appears in a transaction, item Y is also likely to appear in the same transaction. In this study, association rules are used to identify relationships among website pages that are frequently accessed by users. This information can be utilized to understand users' navigation behavior, improve the website structure, and support decision-making in the development of web-based services.

2.4. Debian Web Server

Debian is an open-source Linux operating system distribution that is widely used as a server platform due to its stability, security, and extensive community support. It is developed and maintained by a global community that continuously provides updates and improvements, enabling it to support a wide range of network and server services. As a server operating system, Debian offers numerous software packages that can be used to deploy web services, database systems, file services, and other network applications. One of Debian's main advantages is its ability to operate reliably over long periods while requiring relatively low system resources. In this study, Debian 12 is used as the operating system for running the Apache Web Server and WordPress. Apache functions as the web server that handles user requests for website pages, while WordPress serves as the Content Management System (CMS) for managing website content. Every user activity on the website is recorded by Apache in the *access.log* file, which serves as the primary data source for this research. The use of Debian as the server platform is based on its ease of configuration, system stability, and compatibility with the supporting software required for data collection and log analysis. Therefore, Debian provides a suitable environment for implementing the Apriori algorithm to analyze user access patterns on the website.

2.5. Apache Access Log

Apache Access Log is a log file automatically generated by the Apache Web Server to record every website access performed by users. This log file contains information about the requests sent by users and the responses returned by the server. The data stored in the access log can be utilized for system monitoring, website traffic analysis, and user behavior analysis.
Whenever a user accesses a website page, Apache records specific information in the *access.log* file. This information includes the user's IP address, access time, request method, requested page, HTTP status code, and the size of the data returned by the server.
In this study, the website consists of several main pages, namely Home, Products, Promotions, Contact, and About Us. User activities on these pages are recorded by the Apache Web Server in the *access.log* file.
Example of Apache access log data:
192.168.10.2 - - [22/Jun/2026:10:09:05 +0700] "GET / HTTP/1.1" 200 10345 "-" "Mozilla/5.0 (X11; Linux x86_64; rv:109.0) Gecko/20100101 Firefox/115.0"
192.168.10.2 - - [22/Jun/2026:10:09:18 +0700] "GET /wp-includes/blocks/navigation/style.min.css?ver=6.2 HTTP/1.1" 200 2573 "http://192.168.10.2/" "Mozilla/5.0 (X11; Linux x86_64; rv:109.0) Gecko/20100101 Firefox/115.0"
Based on the log data, information such as the user's IP address, access time, requested page, request method, server response status, and data size can be obtained. Before applying the Apriori algorithm, the access log data undergo preprocessing, including filtering, cleaning, and sessionization, and is then transformed into transaction data (Han et al., 2012).

Table 1: Example of Transaction Data After Preprocessing

Transaction ID	Accessed Pages
U1	Home, Products, Promotions
U2	Home, Products
U3	Home, About Us
U4	Products, Promotions
U5	Home, Contact

The transaction data are used as input for the Apriori algorithm to discover frequent itemsets and generate association rules (Agrawal & Srikant, 1994).

2.6. WordPress

WordPress is an open-source Content Management System (CMS) used to build and manage websites in a simple and flexible manner (WordPress Foundation, 2024). It is developed using the PHP programming language and utilizes MySQL or MariaDB as its database management system (The PHP Group, 2024). In this study, WordPress is used to generate user activities, which are subsequently recorded by the Apache Web Server running on the Debian 12 operating system. Every user request is recorded in the *access.log* file and serves as the primary data source for Apriori analysis.

2.7. Apache Access Log

Apache Access Log is a log file that records all user activities on a website handled by the Apache Web Server. The log contains information such as the user's IP address, access time, requested URL, HTTP request method, response status code, data size, referrer, and user agent. In this study, the access log data undergo preprocessing, including data filtering, noise removal, conversion of *page_id* values into page names, and transaction formation before being analyzed.

2.8. Previous Studies

Previous studies have demonstrated that the Apriori algorithm is effective in discovering relationships among items through association rule mining [8]. Furthermore, research in Web Usage Mining has shown that access log data can be utilized to understand user behavior on websites. This study differs from previous research by using Apache access log data collected from a Debian-based web server as the primary data source instead of sales transaction data. The log data are transformed into transaction data before being analyzed using the Apriori algorithm to generate association rules.

2.9. Relationship Among the Research Concepts

This study begins by deploying a Debian-based web server (Debian Project, 2024). The Apache Web Server is then configured to handle user requests (The Apache Software Foundation, 2024). The website is developed using WordPress so that every user activity is automatically recorded in the Apache Access Log (WordPress Foundation, 2024). The log data are subsequently processed through preprocessing steps, including filtering, cleaning, and sessionization, to produce transaction data (Han et al., 2012). The transaction data are then analyzed using the Apriori algorithm to generate frequent itemsets and association rules (Agrawal & Srikant, 1994). The resulting association rules are used to analyze user access patterns and improve the website's navigation structure.

3. Research Methodology

3.1. Research Method

This study employed an experimental method with a quantitative approach. The experimental method was selected because the research focuses on the implementation of the Apriori algorithm to analyze user access patterns based on access log data generated by a Debian web server. The quantitative approach was applied because the data processed consist of numerical values that were analyzed to produce support and confidence values in the data mining process (Larose, 2014). The research process began with the development of the server environment, followed by access log data collection, data preprocessing, implementation of the Apriori algorithm, and analysis of the resulting association rules. The data used in this study were primary data obtained from user activities on the Fikom Computer Store website, which was developed using WordPress and hosted on an Apache Web Server running on the Debian operating system. The collected data were then transformed into a transaction dataset used as input for the Apriori algorithm (Agrawal & Srikant, 1994).

3.2. Research Stages

This research was conducted through several interconnected stages. The first stage was a literature review, which involved collecting references from books, journals, and previous studies related to Web Servers, Apache, WordPress, Data Mining, Web Usage Mining, and the Apriori algorithm (Witten et al., 2017). The next stage involved implementing the web server using Debian as the primary server operating system. During this stage, Apache was installed as the web server, while PHP and MariaDB were installed to support the WordPress platform (Debian Project, 2024; The PHP Group, 2024). The following stage involved collecting access log data generated by the Apache Web Server during user interactions with the website. These access logs served as the primary data source for this research. Subsequently, data preprocessing was performed, including data filtering, removal of irrelevant records, conversion of page IDs into page names, and grouping user activities into transaction sessions. This preprocessing stage is an essential part of Web Usage Mining for

producing a dataset suitable for analysis (Cooley et al., 1999). Finally, the Apriori algorithm was implemented using the Python programming language to calculate support and confidence values and generate association rules.

Table 2: Research Stages

Stage	Description
Literature Review	Collecting references on Web Mining, Data Mining, and the Apriori algorithm
Server Implementation	Installing Debian, Apache, PHP, MariaDB, and WordPress
Data Collection	Collecting data from the Apache access.log file
Data Preprocessing	Filtering, cleaning, and transaction generation
Apriori Implementation	Processing data using Python
Result Analysis	Analyzing support, confidence, and association rules

3.3. Web Server Implementation

The web server in this study was implemented using the Debian operating system because of its stability in server management. Apache was utilized as the primary web server to handle user requests for website pages. In addition, PHP was employed as the server-side programming language to run WordPress, while MariaDB served as the database management system (The Apache Software Foundation, 2024). The website developed in this research was the Fikom Computer Store website, consisting of several main pages, namely Home, Products, Promotions, About Us, and Contact. All user activities on the website were automatically recorded by Apache in the *access.log* file.

3.4. Data Collection

Data collection was carried out by retrieving records from the *access.log* file located in the following directory:

$$/var/log/apache2/access.log$$

The *access.log* file records all user activities while accessing the website, including IP addresses, access time, HTTP methods, requested URLs, HTTP status codes, data size, referrer information, and user agent details (Srivastava et al., 2000). The data were collected during website testing involving various user activities to produce sufficient data variation for analysis.

Table 3: Sample of Collected Data

IP Address	URL Request
192.168.10.3	/?page_id=13
192.168.10.3	/?page_id=17
192.168.10.3	/?page_id=19

3.5. Data Preprocessing

The preprocessing stage is an important step in this research because the access log data are initially in raw format and cannot be directly used for data mining. The preprocessing process consisted of several stages, including data filtering to remove irrelevant records, exclusion of WordPress administrator activities, conversion of page IDs into page names, and grouping user activities into user sessions. After these processes, the data were transformed into transaction format containing the collection of pages accessed by each user within a single session. This stage is an essential component of Web Usage Mining for generating data suitable for mining analysis (Cooley et al., 1999).

Table 4: Transaction Data after Preprocessing

Transaction	Items (Visited Pages)
U1	Contact, Products, Promotions, About Us
U2	Home, Contact, Products, Promotions, About Us
U3	Home, Products, Promotions

The transaction data were then converted into one-hot encoding format before being processed using the Apriori algorithm. The research parameters are presented below.

Table 5: Research Parameters

Parameter	Value
Minimum Support	30%
Minimum Confidence	60%
Algorithm	Apriori
Programming Language	Python 3
Library	Pandas, Mlxtend

3.6. Data Analysis

Data analysis was conducted on the frequent itemsets and association rules generated by the Apriori algorithm. The support value was used to determine the frequency of occurrence of an item within all transactions, while the confidence value was used to measure the strength of the relationship between items in a transaction. The objective of this analysis was to identify user access patterns that could be utilized to improve the website navigation structure (Agrawal & Srikant, 1994).

3.7. Research Instruments

The research instruments consisted of hardware and software used to support the entire research process, from server implementation to data analysis.

Table 6: Hardware Specifications

Component	Specification
Processor	Intel Core i3 or equivalent
RAM	Minimum 8 GB
Storage	Minimum 100 GB

Table 7: Software Specifications

Software	Function
Debian 12	Server Operating System
Apache2	Web Server
PHP	Server-side Programming Language
MariaDB	Database Management System
WordPress	Content Management System (CMS)

4. Results

4.1. Debian Web Server Implementation

The first stage of this research involved implementing a Debian Web Server as the server environment. The Debian operating system was selected because of its high stability, open-source nature, and widespread use as a server operating system in various network environments. After the operating system installation was completed, Apache Web Server, PHP, and MariaDB were installed to support the operation of the website. Subsequently, WordPress was installed as the Content Management System (CMS), enabling the website to be developed more efficiently. The website developed for this research was named Fikom Computer Store, which consists of several main pages, namely Home, Products, Promotions, About Us, and Contact. After all services were successfully configured and running, the website became accessible through a web browser using the server's IP address within the local network. All user activities on the website were automatically recorded by Apache in the *access.log* file, which served as the primary data source for this study.

4.1.1. Home Page

The Home page is the main page of the website and serves as the first page accessed by users when they visit the website.

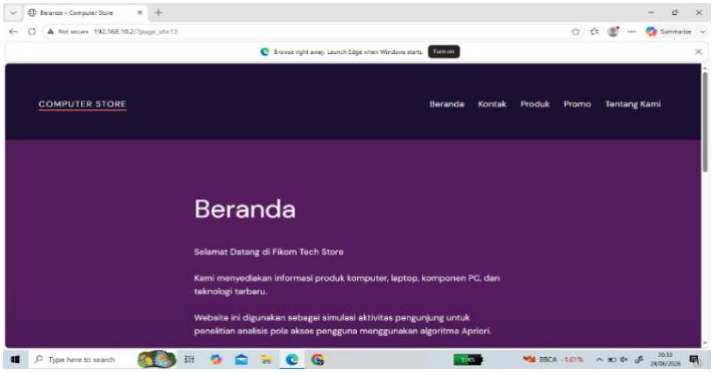

Figure 1: Home Page of the Website

4.1.2. Products Page

The Products page is a page that displays the list of products or services available on the website.

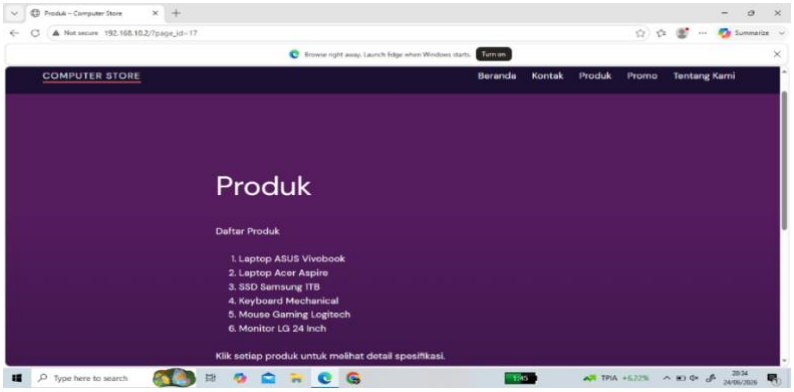

Figure 2: Product Page of the Website

4.1.3. Promotions Page

The Promotions page displays information about various promotional offers, discounts, and promotional programs available on the website.

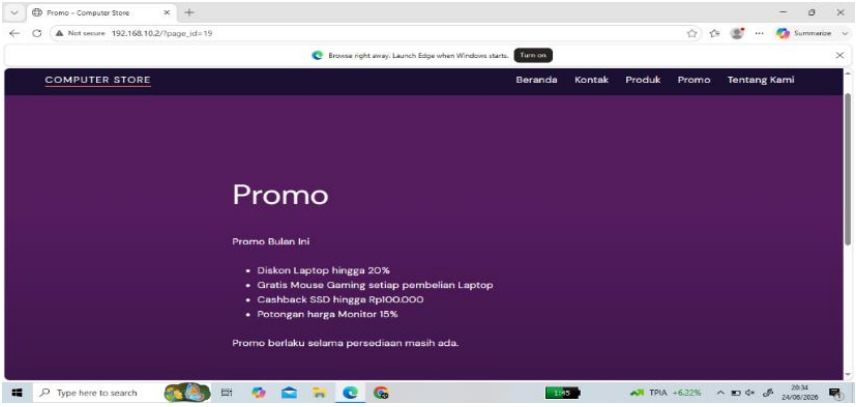

Figure 3: Promotions Page of the Website

4.1.4. About Us Page

The About Us page provides information about the website's profile, its development objectives, and general information about the organization or the administrators responsible for managing the website.

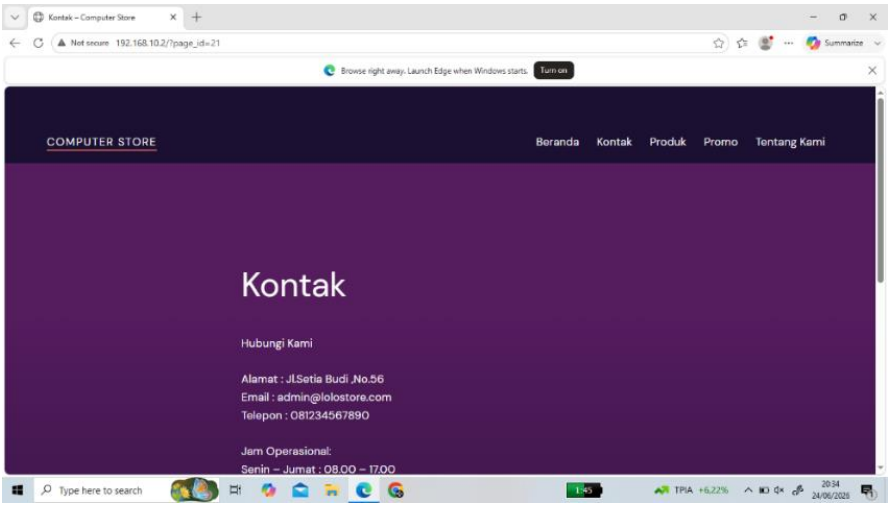

Figure 4: About Us Page of the Website

4.1.5. Contact Page

The Contact page provides information on how users can contact the website administrators.

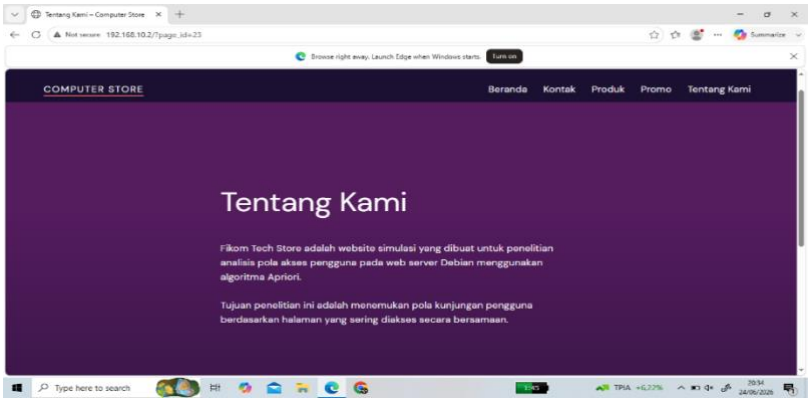

Figure 5: Contact Page of the Website

4.2. Access Log Data Collection

The research data were obtained from the *access.log* file located in the /var/log/apache2/ directory. This file records all user activities while accessing the website, including IP addresses, access time, HTTP methods, requested URLs, HTTP status codes, and browser information. A sample of the log data collected from Apache is shown in Figure 4.2. The data are still in raw format and therefore cannot be directly used for analysis using the Apriori algorithm.

Sample Log Data:

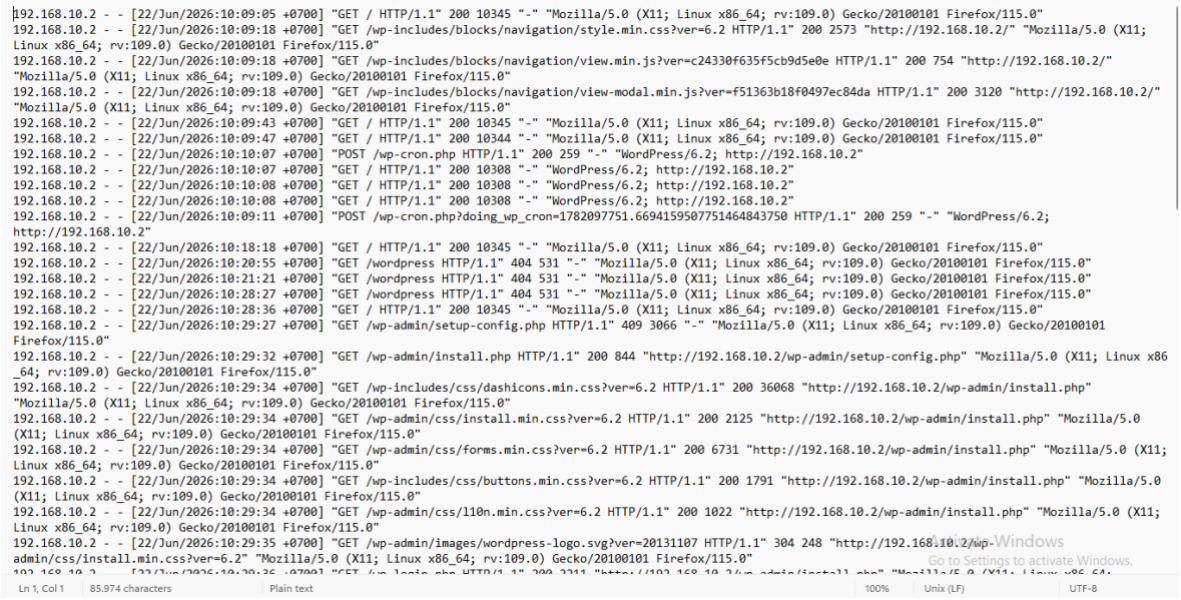

Figure 6: Sample of Apache Access Log Data

4.3. Data Preprocessing

The preprocessing stage was carried out to produce data that were ready for analysis. At this stage, the access log data were filtered by removing irrelevant records, such as favicon requests, WordPress administrator page accesses, CSS files, JavaScript files, and image files. Next, the page_id values were converted into website page names using the WordPress page mapping table. Afterward, the data were grouped based on user activities to generate transaction data, which served as the input for the Apriori algorithm.

Table 8: Page ID Mapping

page_id	Page Name
13	Home
17	Products
19	Promotions
21	Contact
23	About Us

Table 9: Preprocessing Results

Transaction	Item 1	Item 2	Item 3	Item 4	Item 5
U1	Contact	Products	Promotions	About Us	
U2	Home	Contact	Products	Promotions	About Us
U3	Home	Contact	Products	Promotions	About Us
U4	Home	Products	Promotions	About Us	
U5	Home	Products	Promotions	About Us	Contact
U6	Home	Contact			
U7	Home	Contact	Products	About Us	
U8	Products	Contact	Promotions		
U9	Home	Products	Promotions	Contact	About Us
U10	Home	Contact	Products	Promotions	
U11	Contact	About Us	Promotions	Home	
U12	Promotions	Products			
U13	Home	Contact	Products		
U14	Home	Promotions	Products		
U15	About Us	Promotions	Home	Contact	
U16	Promotions	Home	About Us	Contact	
U17	Home	Products	Contact	Promotions	About Us
U18	Home	Products	Contact	Promotions	About Us

4.4. Apriori Algorithm Implementation

The transaction data obtained from the preprocessing stage were subsequently processed using the Apriori algorithm implemented in the Python programming language. The implementation utilized the Pandas and Mlxtend libraries. In this study, the minimum support threshold was set to 30%, while the minimum confidence threshold was set to 60%. The implementation process began by transforming the transaction data into a one-hot encoding format. Subsequently, frequent itemsets that satisfied the minimum support threshold were

identified. Based on the resulting frequent itemsets, association rules were generated using the predefined minimum confidence threshold. The research parameters used in the implementation of the Apriori algorithm are presented in Table.

Table 10: Research Parameters

Parameter	Value
Minimum Support	30%
Minimum Confidence	60%
Algorithm	Apriori
Programming Language	Python

```
===== FREQUENT ITEMSET =====
     support                                          itemsets
0   0.833333                        frozenset({Beranda})
1   0.833333                         frozenset({Kontak})
2   0.777778                         frozenset({Produk})
3   0.833333                          frozenset({Promo})
4   0.666667                  frozenset({Tentang Kami})
5   0.722222               frozenset({Beranda, Kontak})
6   0.611111               frozenset({Beranda, Produk})
7   0.666667                frozenset({Beranda, Promo})
8   0.611111         frozenset({Beranda, Tentang Kami})
9   0.611111                frozenset({Kontak, Produk})
10  0.666667                 frozenset({Promo, Kontak})
11  0.611111          frozenset({Kontak, Tentang Kami})
12  0.666667                 frozenset({Promo, Produk})
13  0.500000          frozenset({Tentang Kami, Produk})
14  0.611111           frozenset({Promo, Tentang Kami})
15  0.500000       frozenset({Beranda, Kontak, Produk})
16  0.555556        frozenset({Beranda, Promo, Kontak})
17  0.555556 frozenset({Beranda, Kontak, Tentang Kami})
```

Figure 7: Apriori Algorithm Implementation

4.5. Frequent Itemset Results

The implementation of the Apriori algorithm produced several frequent itemsets that satisfied the minimum support threshold of 30%. Frequent itemsets represent individual web pages or combinations of web pages that were most frequently accessed together by users during the website testing process. A higher support value indicates that an item or combination of items appears more frequently across all transactions.

Table 11: Frequent Itemset Results (1-Itemset)

Frequent Itemset	Support
Home	83.33%
Promotion	83.33%
Contact	83.33%
Products	77.78%
About Us	66.67%

Based on the support calculation results, the Home, Promotion, and Contact pages achieved the highest support value of 83.33%. This indicates that these three pages were the most frequently accessed by users during the website testing process. Furthermore, the Products page obtained a support value of 77.78%, while the About Us page achieved 66.67%. These findings indicate that all major pages of the website were frequently accessed and satisfied the predefined minimum support threshold.

Furthermore, the Apriori algorithm also generated combinations of two pages (2-itemsets) that frequently appeared together, as presented in Table.

Table 12: Frequent Itemset Results (2-Itemset)

Frequent Itemset	Support
Home, Promotion	66.67%
Products, Promotion	66.67%
Promotion, Contact	66.67%
Home, About Us	61.11%
Products, Contact	61.11%
About Us, Contact	61.11%
Home, Products	61.11%
Promotion, About Us	61.11%

In addition to the 2-itemsets, the Apriori algorithm also generated 3-itemsets, 4-itemsets, and 5-itemsets that satisfied the minimum support threshold of 30%. This indicates that during a single browsing session, users tended to visit multiple pages rather than only one or two pages, resulting in consistent website navigation patterns.

4.6. Association Rule Results

Based on the frequent itemsets obtained, the next step was to generate association rules using a minimum confidence threshold of 60%. Association rules were used to identify relationships among website pages based on users' navigation patterns.

Table 13: Association Rule Results

Rule	Support	Confidence
Home → Products	61.11%	73.33%
Products → Promotion	61.11%	84.62%
Contact → Home	66.67%	85.71%

The association rule results indicate strong relationships among several website pages. High confidence values suggest that when users access a particular page, there is a high probability that they will continue navigating to another related page.

4.7. Discussion

Based on the implementation results of the Apriori algorithm, several user navigation patterns were identified on the Fikom Computer Store website. The frequent itemset analysis revealed that the Home, Promotion, and Contact pages achieved the highest support value of 83.33%, indicating that these pages were the most frequently visited during the website testing process. Meanwhile, the Products page obtained a support value of 77.78%, while the About Us page achieved 66.67%, demonstrating that all primary pages of the website were accessed frequently. These findings suggest that user traffic was relatively well distributed across the website rather than being concentrated on a single page. For page combinations, the frequent itemsets Home–Promotion, Products–Promotion, and Promotion–Contact each obtained a support value of 66.67%, indicating that these page combinations were frequently accessed together within the same transaction. In addition, combinations such as Home–Products, Products–Contact, Home–About Us, About Us–Contact, and Promotion–About Us also satisfied the predefined minimum support threshold. These patterns indicate meaningful relationships among the website pages within the navigation structure [9]. The association rule analysis showed that the rule Products → Promotion achieved a confidence value of 84.62%, indicating that most users who visited the Products page subsequently accessed the Promotion page. Similarly, the rule Contact → Home achieved the highest confidence value of 85.71%, suggesting that users who visited the Contact page tended to return to the Home page. In addition, the rule Home → Products obtained a confidence value of 73.33%, indicating that the Home page serves as the primary entry point for users before navigating to the Products page. Overall, the results demonstrate that the Apriori algorithm successfully identified user navigation patterns based on Apache Access Log data collected from a website hosted on a Debian Web Server. The information obtained from the frequent itemsets and association rules can be utilized to improve the website's navigation structure, optimize menu organization, and determine more effective information placement, thereby enhancing the overall user experience when accessing the website [10].

5. Conclusion

This study successfully implemented the Apriori algorithm to analyze user access patterns on a website hosted on a Debian Web Server. The website was developed using the Apache Web Server and WordPress, while user activities were recorded through Apache access log files, which served as the primary data source for this research. The collected access log data could not be directly analyzed because they contained irrelevant information. Therefore, a preprocessing stage was conducted, including data filtering, removal of administrator page requests and system file accesses, conversion of page_id values into corresponding page names, and construction of transaction data based on user activities. These preprocessing steps produced a dataset suitable for analysis using the Apriori algorithm. The implementation results showed that the Home page was the most frequently accessed page, with a support value of 83.33%, followed by the Contact and Promotion pages, each with a support value of 77.78%, and the Products page with a support value of 72.22%. These findings indicate that these pages served as the primary navigation points for users during website browsing. Furthermore, the association rule analysis generated several relationships among the website pages. One of the strongest rules was Products → Promotion, with a confidence value of 84.62%, indicating that most users who visited the Products page subsequently accessed the Promotion page. The rule Contact → Home achieved a confidence value of 85.71%, while the rule Home → Products obtained a confidence value of 73.33%. These rules describe the navigation behavior of users while accessing the website. Based on the findings, it can be concluded that the Apriori algorithm is effective in identifying relationships among website pages using Apache access log data. The resulting information can serve as a basis for evaluating the website navigation structure, optimizing menu organization, and improving the presentation of information according to user navigation patterns, thereby enhancing the overall quality of the website.

Acknowledgement

The authors would like to express their deepest gratitude to Almighty God for His blessings and guidance throughout the completion of this research. The authors also sincerely thank the supervisor and the Department of Computer Science / Informatics Engineering,

Universitas Katolik Santo Thomas, for their guidance, support, and encouragement during this study, which utilized Debian, Apache, WordPress, and the Apriori algorithm to analyze Apache access log data.

References

[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, 1994.
[2] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.
[3] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed. Hoboken, NJ, USA: Wiley, 2014.
[4] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2017.
[5] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explorations Newsletter*, vol. 1, no. 2, pp. 12–23, 2000.
[6] S. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web," in *Proc. 9th IEEE Int. Conf. Tools with Artificial Intelligence*, 1999.
[7] The Apache Software Foundation, "Apache HTTP Server Documentation," 2024. Available: https://httpd.apache.org/
[8] Debian Project, "Debian Documentation," 2024. Available: https://www.debian.org/
[9] WordPress Foundation, "WordPress Documentation," 2024. Available: https://wordpress.org/
[10] The PHP Group, "PHP Manual," 2024. Available: https://www.php.net/