
OPTIMIZING MARKETING STRATEGIES USING FP-GROWTH AND ASSOCIATION RULE MINING ALGORITHMS IN THE TEXTILE INDUSTRY

Wijaya NG¹, Robby Sukma², Christina Juliane³

STMIK LIKMI, Jawa Barat, Indonesia

wijayavandana@gmail.com¹, Robbysukma91@gmail.com², christina.juliane@likmi.ac.id³

ABSTRACT

This study leverages association rule mining to analyze transaction data from PT. Labda Anugerah Tekstil, a prominent player in the textile industry, to uncover significant purchasing patterns and associations between different fabric types. Utilizing data from January 1, 2022, to December 31, 2023, which includes 7,143 transaction entries, the research applies the FP-Growth algorithm followed by Association Rule Mining to identify and evaluate frequent itemsets and strong association rules within the dataset. The analysis revealed robust associations among fabrics such as Cotton, Linen, Rayon, and Viscose, suggesting substantial opportunities for targeted marketing strategies and inventory management enhancements. The findings indicate that strategically bundling and promoting associated fabrics can drive higher sales volumes and improve customer purchasing experiences. The insights from this study provide actionable strategies for optimizing marketing efforts and inventory management, aiming to enhance sales performance and customer satisfaction in the competitive textile market.

Keywords: Association Rule Mining, FP-Growth Algorithm, Textile Industry Marketing, Customer Purchasing Behavior, Inventory Management Strategies

Corresponding Author: Wijaya NG

E-mail wijayavandana@gmail.com

INTRODUCTION

The textile industry is one of the most critical sectors of the global economy, providing clothing and fabrics for various applications (Tsai et al., 2020). However, the industry faces many challenges, such as high competition, low profit margins, changing customer preferences, and environmental issues (Nugroho & Fadhilah, 2023). To remain relevant and competitive, textile companies must adopt data-driven marketing strategies (Majumdar et al., 2021); (Abbate et al., 2024). PT. Labda Anugerah Tekstil, a key player in the textile industry, continuously strives to leverage the vast data generated from customer interactions and sales transactions to understand the ever-changing market dynamics.

Extensive data analysis has become crucial in identifying patterns and trends in purchasing behavior that are not immediately apparent (Ramkumar et al., 2023). One way to achieve this is by using data mining techniques to analyze customers' transaction data and discover valuable patterns and insights (Soepriyono & Triayudi, 2023). Data mining can help textile companies understand the behavior and preferences of their customers, identify market segments and niches, and design effective promotional campaigns and pricing policies (Shah et al., 2021).

One of the most popular data mining techniques for analyzing transaction data is association rule mining (ARM) (Shaukat et al., 2015), which aims to find rules that describe the relationships between items in a transaction database (Safitri, 2022). For example, an association rule can state that jeans customers are likely to buy T-shirts, too. Such rules can help textile companies recommend products to customers, cross-sell and up-sell products, and increase customer loyalty and retention (Arcos & Hernandez, 2019).

However, traditional ARM methods, such as the Apriori algorithm, could be more efficient and practical for large-scale transaction databases because they generate many candidate item sets and require multiple scans (Gu et al., 2011). To overcome this problem, a more efficient and scalable algorithm, FP-Growth, was proposed by (Jiang and Meng, 2017) (Yin et al., 2018). The FP-Growth algorithm uses a compressed data structure, called a frequent pattern tree (FP-tree), to store the frequent itemsets and their support counts and mines the association rules from the FP-tree without generating candidate itemsets (Ahmed & Nath, 2019); (Pan et al., 2018).

In this study, we apply the FP-Growth algorithm and ARM methods to analyze the transaction data of Labda Anugerah Tekstil, a textile company in Indonesia that produces and sells various types of fabrics and garments. Our main objectives are: 1) To find the frequent itemsets and association rules from the transaction data of Labda Anugerah Tekstil. 2) To evaluate the quality and usefulness of the association rules using various measures, such as support, confidence, lift, and conviction. 3) To provide recommendations and suggestions for optimizing the marketing strategy of Labda Anugerah Tekstil based on the association rules. The rest of the paper is organized as follows. Section 2 reviews the related literature on ARM and FP growth. Section 3 describes the data and methodology used in this study. In section 4, we present and discuss the results of the analysis. In section 5, we conclude the paper and suggest some directions for future research.

METHOD

This study employs quantitative research methods to analyze transaction data from PT. Labda Anugerah Tekstil using data mining techniques. The methodology is structured as follows:

Data Collection

The dataset used in this study consists of transaction records from PT. Labda Anugerah Tekstil spans a specified period. The dataset used in this study comprises fabric and clothes sales records spanning from January 1, 2022, to December 31, 2023. The data, totaling 7,143 rows, includes detailed transaction entries with fields for Date, Customer Name, and quantities of fabrics such as Cotton, Linen, Rami, Rayon, Silk, Tencel, and Viscose. This comprehensive dataset provides an overview of buying patterns and customer preferences within the company's operations.

Data Preparation

The preparation of the dataset involves several crucial steps to ensure the data is suitable for mining:

1. **Data Cleaning:** Handling missing values, removing duplicate entries, and ensuring the consistency of data formats across the dataset.
2. **Data Transformation:** Converting the relevant numerical data into a binominal format where fabric types are represented as binary attributes. This transformation is crucial for the subsequent application of the FP-Growth algorithm.

Data Mining Techniques

FP-Growth Algorithm

The FP-Growth algorithm will efficiently discover frequent itemsets in the dataset. This algorithm is chosen for its efficiency in handling large datasets without candidate generation, significantly reducing the computational overhead compared to traditional apriori-like approaches.

Association Rule Mining

After identifying frequent item sets, association rules will be generated using a minimum confidence threshold. The rules will identify which fabric types are frequently purchased together.

Parameters such as support, confidence, and lift will be calculated for each rule to assess its strength and relevance.

Analysis

The analysis will focus on interpreting the association rules to understand the co-purchasing patterns among fabric types. The results will draw insights into consumer behavior and preferences, identifying potential cross-selling opportunities and effective promotion combinations.

Validation

To ensure the findings' reliability, the generated rules will be validated against known marketing and sales strategies PT employs. Labda Anugerah Tekstil. The impact of these rules on sales performance will be analyzed to measure the effectiveness of data-driven decision-making in the textile industry.

Tools

The study will utilize RapidMiner for data processing and analysis due to its robust data mining capabilities and user-friendly interface for handling complex datasets and performing advanced analytical processes. Below is the flowchart for this research:

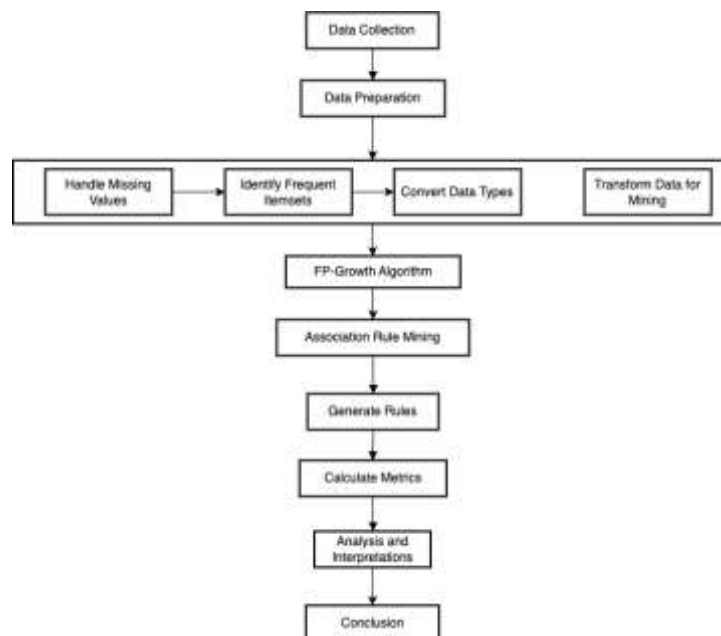


Figure 1. Research Flowchart

Our analysis's initial stage begins with providing raw transaction data from PT. Labda Anugerah Tekstil. This dataset includes comprehensive details of each transaction, encompassing attributes such as Date, Customer Name, and quantities of fabric types like Cotton, Linen, Rayon, Silk, Tencel, and Viscose. Here, I will display the raw data table to provide a snapshot of the information as it was initially recorded. This table will include several rows of data, each representing a unique transaction, filled with all original details and including any inconsistencies or anomalies present in the dataset.

Table 1. Fabric Sales 2022-2023 Labda Anugerah Textile

No	Customer	Fabric	Date	Rayon	Cotton	Linen	Viscose	Silk	Tencel	Rami
1	TIO	Rayon	Jan-22	1,25	-	-	-	-	-	-
2	TIO	Cotton	Jan-22	-	12,00	-	-	-	-	-
3	TIO	Rayon	Jan-22	32,00	-	-	-	-	-	-
4	TIO	Cotton	Jan-22	-	106,00	-	-	-	-	-
5	SYA	Silk	Jan-22	-	-	-	-	52,82	-	-
6	SYA	Rayon	Jan-22	160,00	-	-	-	-	-	-
7	SYA	Rayon	Jan-22	120,00	-	-	-	-	-	-
8	SYA	Rayon	Jan-22	160,00	-	-	-	-	-	-
9	SYA	Rayon	Jan-22	123,00	-	-	-	-	-	-
10	SYA	Linen	Jan-22	-	-	2,00	-	-	-	-
11	SYA	Linen	Jan-22	-	-	2,00	-	-	-	-
7143 SIM Viscose	Dec-23	-	-	-	19,60	-	-	-

The raw data often contains issues that could skew the analysis, such as missing values, duplicate records, and incorrect data entries. To address these problems, we apply a series of data-cleaning procedures:

- 1. Handling Missing Values:** We identify and address missing entries in the dataset. Depending on the nature of the missing data, we either fill these gaps with estimated values (using methods such as the mean or median of the column) or remove rows entirely if the missing data comprises critical information that cannot be reliably estimated.
- 2. Removing Duplicates:** We eliminate duplicate entries to ensure each transaction is unique. This is crucial to prevent bias from the same transaction being recorded multiple times.
- 3. Correcting Data Anomalies:** Any anomalies or outliers that do not make sense within the context of the data (such as the unrealistic confidence level mentioned earlier) are corrected or removed after a thorough investigation to determine their cause.

Following the data cleaning steps, I will display the cleaned data table. This table will be free from duplicates, filled with missing values, and corrected anomalies. It will be ready for further analysis. This version of the dataset is what we use to conduct our association rule mining.

Table 2. Association Rule Mining Dataset

No.	Customer Name	Cotton	Linen	Rami	Rayon	Silk	Tencel	Viscose
1	ATI	1	0	0	0	0	0	0
2	SYA	0	1	0	1	1	0	0
3	TIO	1	0	0	1	0	0	0
4	AJI	0	0	0	1	0	0	0
5	ATI	1	0	0	0	0	0	0
649 WYA 1 0 0 1 0 0 0

After the initial data cleaning phase, we transformed the dataset to align it for practical mining of association rules. This transformation involved two critical steps: data pivoting and data binarization.

Data Pivoting: To structure the data more effectively for analysis, we performed a pivoting operation based on the transaction date. This process reorganized the dataset so that each row represents a customer's daily transactions. The columns of this pivoted table correspond to various types of fabrics, such as Cotton, Linen, Rayon, Silk, Tencel, and Viscose. The resulting table

aggregates the data such that each entry indicates all the fabric types purchased by a customer on a specific date.

Data Binarization: Following the pivoting of data, the next step was to address the issue of missing values—instances where no purchase quantity was recorded. In this context, missing values were replaced with 0, indicating that the customer did not purchase the fabric on that date. Conversely, any positive quantity indicating a purchase was replaced with 1. This binarization transforms the dataset into a binary format, where 1 signifies a fabric's presence (purchase), and 0 indicates its absence (no purchase). This step is crucial for the subsequent application of data mining algorithms as it simplifies the dataset, making it amenable to the algorithms' requirements.

Implementation of Data Mining Algorithms: With the data now properly formatted—grouped by customer and date and binarized to indicate the presence or absence of fabric types—we are poised to apply the FP-Growth algorithm. The FP-Growth algorithm will allow us to efficiently discover frequent itemsets within the dataset, which are groups of items (fabrics, in this case) that occur together frequently in transactions. Following identifying these frequent itemsets, we will employ Association Rule Mining to explore and identify strict rules where certain fabrics in a transaction imply the presence of others. These rules are evaluated based on metrics such as support, confidence, and lift, which provide insights into the strength and reliability of the associations.

The analysis of association rules within the transaction data of PT. Labda Anugerah Tekstil has yielded significant insights into the co-purchasing patterns of different fabric types. The study identified strong association rules among fabrics such as Cotton, Linen, Rayon, and Viscose.

Table 3. The Transaction Data

No	Premise Conviction	Premise Items			Conclusion		Conclusion Items			Confidence	
		Gain	Laplace Lift	Ps			Total Support				
1	Cotton, Linen	2	Rayon	1	0,500	1,162	-0,227	0,934	1,193	0,012	0,012
2	Cotton, Silk	2	Linen	1	0,500	1,495	-0,180	0,946	1,979	0,030	0,060
3	Rayon, Linen	2	Cotton, Viscose	2	0,500	1,716	-0,143	0,956	3,527	0,034	0,048
4	Cotton, Silk	2	Viscose	1	0,513	1,562	-0,179	0,948	2,147	0,033	0,062
5	Rayon, Viscose	2	Cotton, Linen	2	0,517	1,757	-0,137	0,959	3,422	0,034	0,048
6	Rayon, Viscose	2	Cotton, Silk	2	0,517	1,820	-0,137	0,959	4,299	0,037	0,048
7	Cotton, Viscose	2	Rayon	1	0,522	1,215	-0,210	0,941	1,245	0,015	0,074
8	Cotton, Silk	2	Rayon	1	0,538	1,259	-0,176	0,950	1,285	0,014	0,065
9	Rayon, Linen	2	Viscose	1	0,548	1,685	-0,139	0,961	2,296	0,030	0,052
10	Linen, Viscose	2	Cotton, Rayon	2	0,554	1,809	-0,125	0,965	2,874	0,031	0,048
11	Rayon, Silk	2	Linen	1	0,566	1,722	-0,117	0,967	2,240	0,026	0,046
12	Rayon, Viscose	2	Linen	1	0,567	1,725	-0,133	0,963	2,242	0,029	0,052
13	Rayon, Viscose	2	Silk	1	0,583	1,938	-0,131	0,965	3,029	0,036	0,054

No	Premise Conviction	Premise Items			Conclusion		Conclusion Items			Confidence	
		Gain	Laplace Lift	Ps			Total Support				
14	Rayon, Silk	2	Cotton, Viscose	2	0,585	2,068	-0,116	0,969	4,126	0,036	0,048
15	Viscose	1	Cotton	1	0,594	1,054	-0,336	0,922	1,038	0,005	0,142
16	Linen	1	Cotton	1	0,598	1,064	-0,354	0,919	1,045	0,007	0,151
17	Linen Slik	2	Rayon	1	0,600	1,452	-0,108	0,971	1,432	0,014	0,046
18	Linen, Viscose	2	Rayon	1	0,607	1,479	-0,120	0,969	1,449	0,016	0,120
19	Silk	1	Cotton	1	624,000	1,139	-0,265	0,939	1,092	0,010	0,120
20	Cotton, Rayon, Linen	3	Viscose	1	0,633	2,072	-0,103	0,974	2,649	0,030	0,048
21	Cotton, Rayon, Viscose	3	Linen	1	0,646	2,110	-0,100	0,976	2,556	0,029	0,048
22	Cotton, Rayon, Viscose	3	Slik	1	0,646	2,280	-0,100	0,976	3,353	0,034	0,048
23	Viscose, Silk	2	Cotton, Rayon	2	0,660	2,372	-0,097	0,977	3,425	0,034	0,048
24	Rayon, Silk	2	Viscose	1	0,660	2,241	-0,109	0,974	2,765	0,034	0,054
25	Cotton, Linen, Viscose	3	Rayon	1	0,689	1,867	-0,091	0,980	1,644	0,019	0,048
26	Cotton, Rayon, Silk	3	Viscose	1	0,738	2,906	-0,082	0,984	3,090	0,032	0,048
27	Viscose, Silk	2	Rayon	1	0,745	2,275	-0,091	0,983	1,777	0,024	0,054
28	Cotton, Viscose, Silk	3	Rayon	1	775,000	2,582	-0,076	0,987	1,849	0,022	0,048
29	Linen, Silk	2	Cotton	1	0,780	1,947	-0,094	0,984	1,364	0,016	0,060
30	Rayon, Linen	2	Cotton	1	0,790	2,043	-0,116	0,116	0,982	1,383	0,021
31	Rayon, Silk	2	Cotton	1	0,792	2,064	-0,099	0,984	1,386	0,018	0,065
32	Rayon, Viscose	2	Cotton	1	0,800	2,142	-0,111	0,983	1,399	0,021	0,074
33	Linen, Viscose	2	Cotton	1	0,804	2,181	-0,103	0,984	1,406	0,020	0,069
34	Viscose, Silk	2	Cotton	1	0,851	2,876	-0,083	0,990	1,489	0,020	0,062
35	Rayon, Viscose, Silk	3	Cotton	1	0,886	3,748	-0,060	0,994	1,549	0,017	0,048
36	Rayon, Linen, Viscose	3	Cotton	1	0,912	4,855	-0,057	0,996	1,595	0,018	0,048

Key Findings:

High Confidence and Lift Values: Several rules demonstrated high confidence and lift values, indicating solid relationships. For example, the rule involving Cotton, Linen, and Rayon concluded

with Viscose (Confidence: 0.689, Lift: 1.867) suggests a strong likelihood that customers purchasing the first three fabrics are also likely to purchase Viscose.

Diverse Fabric Combinations: The study highlighted the frequent combination of traditional and modern fabrics, such as the pairing of Cotton and Silk with modern synthetic fibers like Viscose and Rayon. This suggests a blending of traditional and contemporary fashion trends among the customers.

Negative Gain Values: Several rules exhibited negative gain, indicating that the occurrence of the conclusion is less than expected under independence, such as the rule (Cotton, Silk → Linen), which showed a gain of -0.180. This could suggest a conditional dependency among these items, where purchasing one may discourage purchasing another in the absence of a third item.

We will first present several graphs before diving into the detailed discussion of our findings from the association rule mining performed on PT Labda Anugerah Tekstil's transaction data. These visualizations are designed to illustrate the relationships and patterns identified through our analysis, providing a clear visual context for the subsequent detailed discussion.



Figure 2. Cotton Graph

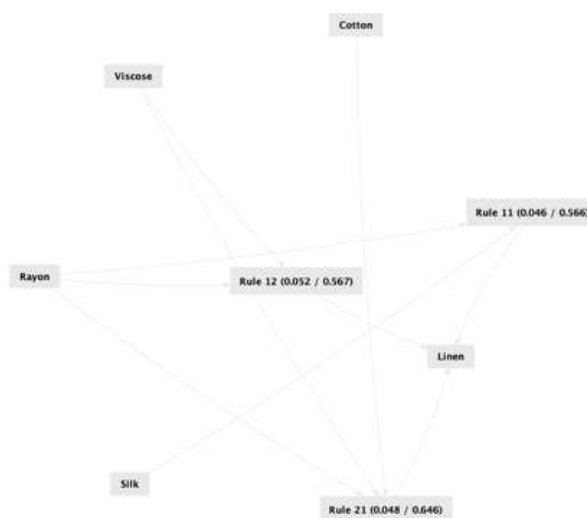


Figure 3. Linen Graph



Figure 4. Rayon Graph



Figure 5. Viscose Graph

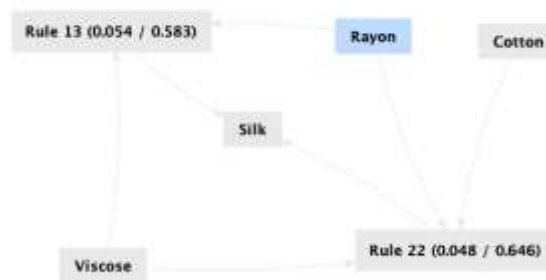


Figure 6. Silk Graph

Each graph represents a network diagram where nodes correspond to fabric types such as Cotton, Linen, Rayon, Silk, and Viscose. Edges between these nodes represent the association rules derived from the data, with the rules' strength and confidence indicated on the connecting lines. Specific metrics such as support and confidence values are displayed alongside each connection. These metrics quantify the strength and reliability of each association rule, aiding in the visual interpretation of how frequently and strongly different fabrics are purchased together. The layout of these diagrams helps identify clusters of fabrics that frequently appear together in transactions.

This clustering provides initial insights into potential customer purchasing patterns and preferences, which are explored in greater depth in the discussion.

The results of the association rule mining on PT will be discussed. Labda Anugerah Tekstil's transaction data brings several strategic insights that could significantly affect the company's marketing strategy. The analysis revealed strong associations between specific groups of fabrics, which suggests the potential for targeted promotional activities or bundled offers to boost sales volumes effectively. For example, promoting Viscose alongside Cotton, Linen, and Rayon could capitalize on their strong association, encouraging customers to purchase these fabrics together.

Regarding inventory management, the findings provide valuable insights that can aid in optimizing stock levels by ensuring that fabric types frequently purchased together are well stocked and placed adjacently PT. Labda Anugerah Tekstil can facilitate cross-selling, which may lead to increased sales. This strategic placement could also enhance the shopping experience, making it easier for customers to find and purchase complementary fabrics.

Additionally, the identified patterns from the data can assist in segmenting customers based on their purchasing preferences. This segmentation can be leveraged for personalized marketing, where promotions and communications are tailored to meet different customer segments' specific needs and preferences. Such targeted marketing efforts could lead to higher conversion rates and customer loyalty.

CONCLUSION

The association rules are derived from PT. Labda Anugerah Tekstil's sales data provide valuable insights into its customers' purchasing habits. The study not only aids in understanding current market dynamics but also offers actionable strategies for enhancing marketing efforts. Leveraging advanced data mining techniques will be crucial for sustaining competitive advantage in the evolving textile market. The implications of the results of this study can be used to design marketing campaigns that are more targeted and efficient, as well as help companies in managing inventory better. By understanding customer buying patterns, companies can identify popular products and areas that need innovation, so as to help new product development. In addition, purchase data provides insights that can be used to set more competitive prices. With the rules of association, companies can also identify different market segments and adjust marketing strategies according to the characteristics of each segment, increasing the relevance and effectiveness of the overall campaign.

REFERENCES

- Abbate, S., Centobelli, P., Cerchione, R., Nadeem, S. P., & Riccio, E. (2024). Sustainability trends and gaps in the textile, apparel and fashion industries. *Environment, Development and Sustainability*, 26(2), 2837–2864.
- Ahmed, S. A., & Nath, B. (2019). Modified fp-growth: an efficient frequent pattern mining approach from fp-tree. *Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17-20, 2019, Proceedings, Part I*, 47–55.
- Arcos, J. R. D., & Hernandez, A. A. (2019). Analyzing online transaction data using association rule mining: Misumi philippines market basket analysis. *Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City*, 45–49.
- Gu, J., Wang, B., Zhang, F., Wang, W., & Gao, M. (2011). *An Improved Apriori Algorithm BT - Applied Informatics and Communication* (D. Zeng (ed.); pp. 127–133). Springer Berlin Heidelberg.
- Jiang, H., & Meng, H. (2017). A parallel FP-growth algorithm based on GPU. *2017 IEEE 14th International Conference on E-Business Engineering (ICEBE)*, 97–102.
- Majumdar, A., Garg, H., & Jain, R. (2021). Managing the barriers of Industry 4.0 adoption and implementation in textile and clothing industry: Interpretive structural model and triple helix framework. *Computers in Industry*, 125, 103372.
- Nugroho, A., & Fadhilah, M. (2023). Customer-Centric Strategy Dalam Menghadapi Persaingan Perusahaan Jasa Konstruksi. *Jurnal Teknologi Dan Manajemen Industri Terapan*, 2(4), 316–325.
- Pan, Z., Liu, P., & Yi, J. (2018). An improved FP-tree algorithm for mining maximal frequent patterns. *2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 309–312.
- Ramkumar, A., Kulkarni, P., Obaid, A. J., Abdulbaqi, A. S., & Yakin, A. Al. (2023). Big data analytics and its application in E-commerce. *AIP Conference Proceedings*, 2736(1).
- Safitri, N. (2022). Penggunaan Algoritma Apriori Dalam Penerapan Data Mining Untuk Analisis Pola Pembelian Pelanggan (Studi Kasus: Toko Diengva Bandar Jaya). *Jurnal Portal Data*, 2(1).
- Shah, S. M., Lütjen, M., & Freitag, M. (2021). Text mining for supply chain risk management in the apparel industry. *Applied Sciences*, 11(5), 2323.
- Shaukat, K., Zaheer, S., & Nawaz, I. (2015). Association rule mining: An application perspective. *International Journal of Computer Science and Innovation*, 2015(1), 29–38.
- Soepriyono, G., & Triayudi, A. (2023). Implementasi Data Mining dengan Algoritma Apriori dalam Menentukan Pola Pembelian Aksesoris Laptop. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(4), 2087–2096.
- Tsai, H. T., Ho, T. H., & Wang, C.-N. (2020). Productivity evaluation of Asia textile industry. *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 620–624.
- Yin, M., Wang, W., Liu, Y., & Jiang, D. (2018). An improvement of FP-Growth association rule mining algorithm based on adjacency table. *MATEC Web of Conferences*, 189, 10012.



© 2024 by the authors. It was submitted for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).