p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 8 August 2024 http://jist.publikasiindonesia.id/

Clothing Product Recommendations Using the FP-Growth Algorithm in Siny.CO Stores

Duny Muhamad Firmansyah[1*], Sunjana[2]
Universitas Widyatama, Indonesia
Email: mfduny@gmail.com[1*], dm.sunjana@gmail.com[2]

*Correspondence

Keywords: association; clothing sales; data mining; fp-growth; rapidminer.

ABSTRACT

In this modern era, there is so much competition in the business world, especially in the sales industry, that requires shop owners to find a strategy that can increase sales and marketing of the products they sell, one of which is by utilizing clothing sales transaction data using data mining. Data Mining is an iterative and interactive process to find new patterns or models that can be generalized for the future, useful and understandable in a very large database (massive database). In these conditions, good data processing techniques are needed, one of which is data mining techniques. One thing that can be used in this technique is to use the Association rule method using the FP-Growth algorithm, which is an algorithm that produces frequent itemsets which will later be used to determine recommendations for clothing products for the needs of the Siny.co Store. This research uses the RapidMiner Studio application to help process transaction data. This research method uses a minimum support of 20% and a minimum confidence of 80%, thereby creating 9 association rules, where these rules determine several products for recommendation at the Siny.Co Store, including Sashi, Canna, Alice, Tartan, Aruna, Cassandra, Nala, Lalita, and Acio.

Introduction

The many competitions in the business world, especially in the sales industry, require developers to find a strategy that can increase sales and marketing of the products sold, one of which is the use of sales transaction data (Kumar & Dubey, 2023). With daily sales activities, the data will increase over time. The data not only functions as an archive for the company, it can also be used and processed into useful information for increasing sales and product promotion (Allam & Dhunny, 2019).

From the source of clothing store sales data, it shows that the demand for clothing is increasing. This is what is used as the basis for processing Data Mining in clothing store sales. To manage this data, a method is needed that can be used to dig up information from the data. This method is known as Data Mining. Data mining is an iterative and

interactive process of finding new patterns or models in massive databases that are useful, understandable, and generalizable for the future (Sari, Muhammad Syahril, Kom, & Suharsil, 2021). Data mining involves searching large databases for trends and patterns of interest in order to make future decisions (Saura, 2021). These patterns can be recognized with specialized tools that can provide useful and insightful data analysis, and can then be explored in more detail, perhaps using other decision support tools (Wardani, 2020).

Previous research on inventory systems using the FP-Growth Algorithm concluded that data mining techniques with the FP-Growth Algorithm can be implemented in the spice product inventory system (Kana, Ramadhan, & Mahyuni, 2022). With an application based on information technology, a method that can increase knowledge to provide goods by providing advice to companies on the relationship of consumers to an item purchased by consumers can be calculated using the FP-Growth Algorithm technique (Salu, Michael, Padang, & Adda, 2022).

The FP-Growth algorithm includes a type of association rule on Data Mining, the FP-Growth Algorithm which aims to find frequent itemsets executed on a set of datam (Wu & Zhang, 2023). FP-Growth analysis defines a process to find all FP-Growth rules that meet the minimum requirements for support and the minimum requirements for confidence (Shawkat, Badawi, El-ghamrawy, Arnous, & El-desoky, 2022). In this study, the FP-Growth algorithm will be used for an association approach, so that the right product recommendations will be found (Brous, Janssen, & Herder, 2020).

Digging association rules is a procedure to find relationships between items in a dataset. Start by looking for frequent itemsets, which are the combinations that most often occur in an itemset and must meet the minimum support (demand). Itemset mining that often arises from large transactional databases is one of the most challenging problems in data mining, in many real-world scenarios, data is not extracted from a single data source but from distributed and heterogeneous data. The knowledge found is expected to help better business operations (Merliani, Khoerida, Widiawati, Triana, & Subarkah, 2022). In the data mining method, association rules are one of the most popular. However, digging up information patterns that use association rules often results in very large individual patterns, thus leaving the analyst to complete the task with all the rules and find the one of interest (Hasan, Aziz, & Nofendri, 2023).

Based on research that has been conducted previously, including research entitled "Application of the Apriori Algorithm in Sales Transactions for Food and Drink Menu Recommendations". In this research, the Apriori method was used to find out the menu at Warung Tenda to be used as a menu package with 50 sales transaction data for a total of 10 items, 11 rule associations were formed which could use these results as reference material to be added as a menu package to the Warung Tenda menu list (Hilman, 2022).

Research Methods

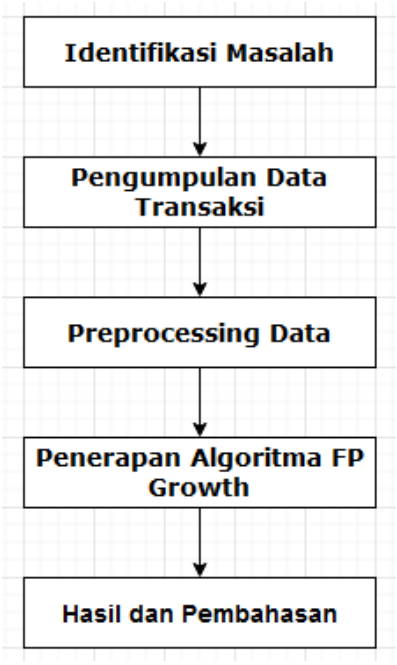


Figure 1. Research Methods

In conducting this research in order to get the appropriate results as desired, the research method that has been determined in accordance with the algorithm in this study is used. The following are the steps used in this research object using the FP-Growth algorithm that has been adjusted, as shown in figure 1 above. The explanation of each of the stages above is as follows:

Problem Identification

The problem identified in this study is to get recommendations related to clothing products in the Siny.Co store to produce maximum product sales as well as maximize product promotion by determining the best product recommendations.

Transaction Data Collection

The data collection stage is the second stage in doing this research task. The data of this research was collected from transaction data on Siny.Co stores. The data comes from transactions that have occurred, namely from January to November in 2022. In collecting data using certain techniques or methods in the process. The following is an explanation of data collection at the Siny.Co Shop:

study book

It is carried out with the aim of finding out what method will be used to solve the problem to be researched, as well as obtaining a strong reference basis in applying a method that will be used in this Research Project, namely by studying articles and journals related to the problem of Clothing Product Recommendations in Siny.Co Stores using the FP-Growth algorithm.

Observation

The observation technique used is direct observation of the object being studied, such as making direct observations at Siny.Co stores. Observation activities are carried out in the circulation room to see transaction activities and products of goods sold.

Preprocessing Data

This stage includes the data processing process using a predetermined method, namely the Association Rule Mining method using the FP - Growth Algorithm. The process of processing sales transaction data is as follows:

1. Selection

At this stage, transaction data that contains information on the item number and item code in the transaction data is not included in the dataset to be processed. In addition, the price of purchased goods is not used in this study and is not included in the FP-Growth process.

2. Cleaning

At this stage, the process of correcting incomplete or blank data is carried out because the transaction is recorded in the exel. The handling for this problem is to add/fill in the blank data so that the dataset is complete.

3. Transformation

In this transformation process, the data is transformed or combined into a format that is suitable for processing in data mining. Transaction data will be made into tabulation data in a boolean table presented in binary form, where the number 1 (one) indicates the existence of a transaction and 0 (zero) indicates the absence of a transaction.

Application of FP Growth Algorithm

The FP-Growth algorithm is carried out using the RapidMiner Studio tool application, the minimum support and confidence is determined by comparing support and confidence values from the highest to the lowest so that effective support and confidence values are obtained. The step to carry out the FP - Growth algorithm by building the data structure used is Tree or commonly known as FP - Tree. The FP-Growth algorithm is divided into three main stages, namely:

a. The conditional pattern base
b. FP-Tree conditional generation stage
c. Frequent itemset search stage
d. Results and Discussion

The results and discussions in the research are in the form of information generated from the data mining process. The information generated is in the form of rules for purchasing items that are often purchased by consumers at the same time, from this point store owners can Siny.Co take a promotional strategy to find out product recommendations for customers and store owners can find out which products can be used as promotions for the future.

Results and Discussion

Data Selection

At this stage, the data used for the research will be selected to match the title of the research to be researched. In this study, secondary data is used, namely transaction data at Siny, Co stores with attributes consisting of dates, customer names, item names, number of goods, and item prices.

Table 1
Clothing Store Transaction Data Before Pre-processing

Date	Customer Name	Item Name	Sum	Price
01/01/22	Ridwan	Tartan	2 pcs	IDR 200.000
02/01/22	Fitri	Sashi	1 pcs	IDR 125.000
03/01/22	Citra	Canna	1 pcs	IDR 125.000
04/01/22	Salman	Tartan	1 pcs	IDR 100.000
05/01/22	Fatimah	Alice	2 pcs	IDR 300.000

Table 1 shows the data of clothing sales transactions before the pre-pcessing stage. Where the data is still intact when data is still being collected at the store Siny.Co each number in the quantity attribute shows the number of clothes purchased.

Data Prepocessing/Cleaning

At this stage, the focus is on the cleaning process. The cleaning process includes transaction data on Siny.co stores in 2022, removing data duplication and several unnecessary attributes, including dates, customer names and prices. The removal of the attribute was carried out because it was considered that it would not affect the rules of association. Then, in this process, changes are made to the number of clothes purchased replaced with 1 and the number of clothes not purchased is replaced with the number 0.

Tabel 2
Data Transaksi Toko Pakaian Setelah Pre – Processing

Customer	Tartan	Sashi	Alice	Canna
1	0	0	0	0
2	0	0	1	0
3	0	0	1	0
4	0	0	1	0
5	0	0	2	0

Transformation

At this stage, data transformation is carried out by providing initialization to attributes that are adjusted to the type needed in the Fp-Growth algorithm. In this transformation process, initialization is carried out on the attributes of the customer data, which can be seen in Figure 2 below.

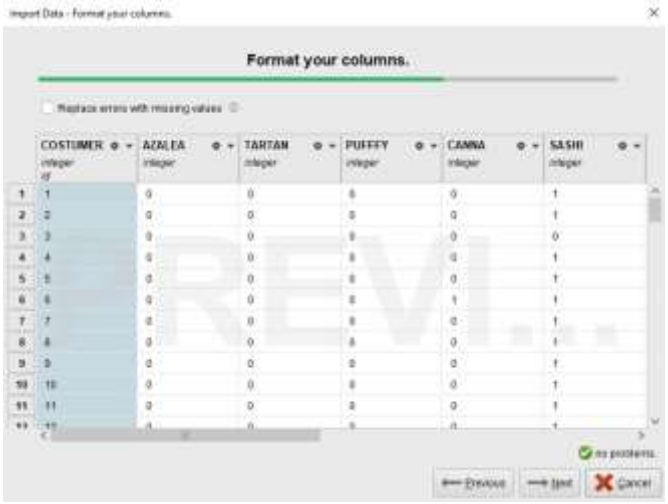


Figure 2. Data Transformation

Application of FP–Growth Algorithm

In this study, the Fp-Growth algorithm is used in clothing sales transactions at Siny.co stores shown in the figure below.

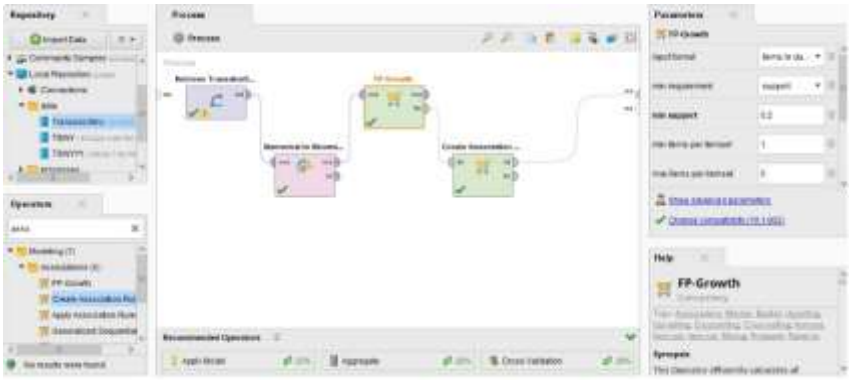


Figure 3. Application of the Fp-Growth Algorithm

Figure 3 above shows the processing of the sales transaction dataset using the RapidMiner application. In this data processing using several operators, including:

1. Clothing sales transaction data retrive

The retrive operator is used to enter clothing sales transaction data after going through the pre-processing process to be subsequently processed by the RapidMiner application.

2. Numerical to binominal

This numerical operator serves to convert the type of numeric attribute to binomial. Binominals are attributes that only have two values, namely true or false.

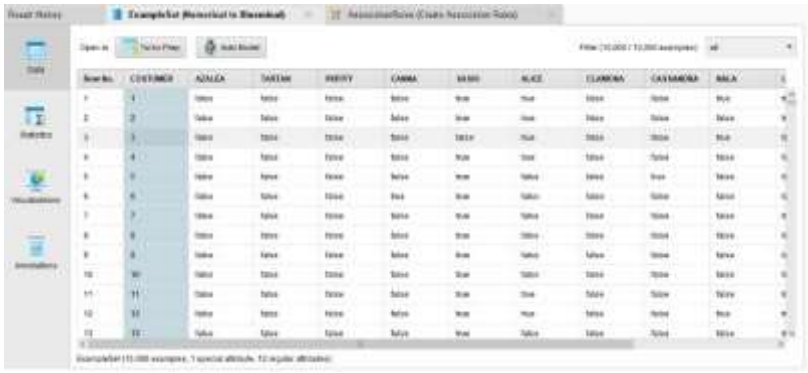


Figure 4. Numerical to binominal results

Figure 4 above shows the results of the numerical to binominal process, where the attributes change to "*true*" and "*false*". A true value indicates a transaction while a *false* value indicates no transaction.

FP-Growth

This Fp-Growth operator is used to identify the *frequency of the itemset* that will be used by the operator later for the association process by determining the *support value* of the sales transaction data at the Siny.co store. Below you can see the results of the *frequency itemset* on clothing sales transactions with a *support value* of 20%.

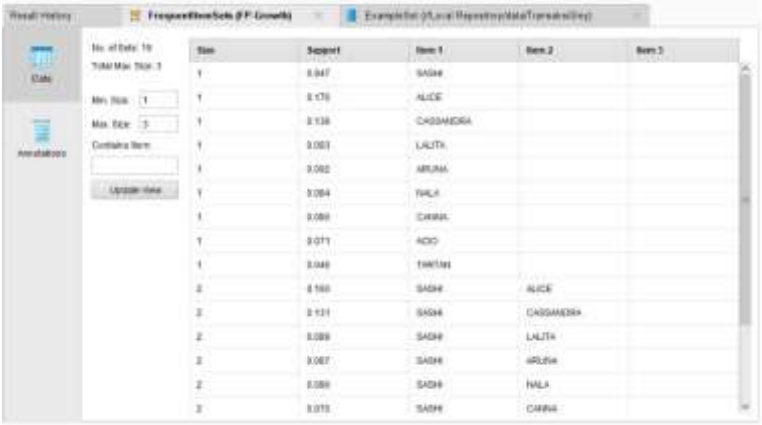


Figure 5. Frekuent Itemset Results

Create Association Rule

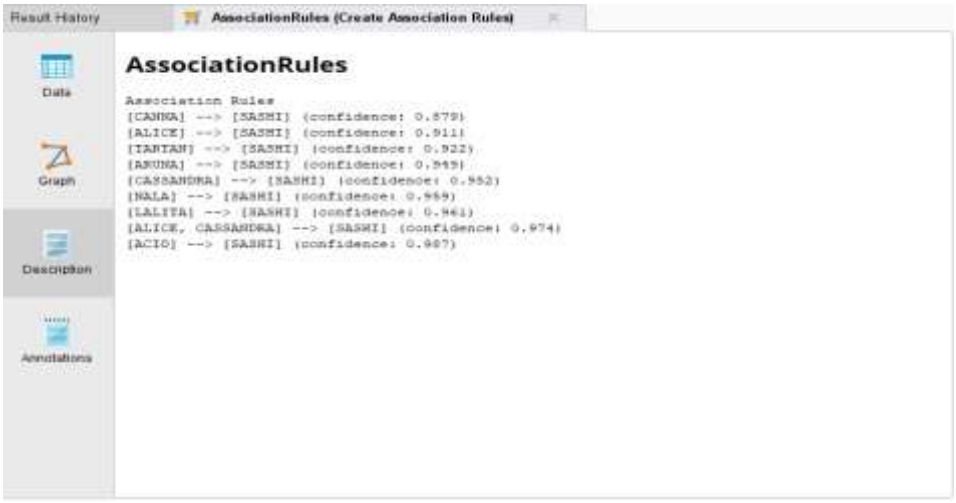
This operator association rule is used to establish associative law by setting a minimum confidence value of an item or itemset from a clothing sales transaction at a Siny.co store.

Result

The results of the Association rule used in this study to determine the pattern of clothing sales transactions in Siny.co stores using the RapidMiner application with a minimum support of 20% and a minimum confidence of 80% resulted in 9 association rules.

Gambar 6. Graph Association Rules

In Figure 6 above is the result of the application of the Fp-Growth algorithm method to clothing sales transactions at Siny.co stores in the form of graphs. Each rule is based on the direction of the arrow.



Gambar 7. Description Association Rules

In Figure 7 above is the result of the application of the Fp-Growth algorithm method to clothing sales transactions at Siny.co stores in the form of descriptions. Where there are 9 rules that are eligible to build association rules on clothing sales transactions at Siny.co stores and are used as the final rules in the dataset analysis. The 9 rules can be explained as follows:

1. If you buy Canna products, then the possibility of buying Sashi products is also with a confidence value of 0.879.
2. If you buy Alice products, then the possibility of buying Sashi products is also with a confidence value of 0.911.
3. If you buy Tartan products, then the possibility of buying Sashi products is also with a confidence value of 0.922.
4. If you buy Aruna products, then the possibility of buying Sashi products is also with a confidence value of 0.949.

5. If you buy Cassandra products, then the possibility of buying Sashi products is also with a confidence value of 0.952.
6. If you buy a Nala product, then the possibility of buying a Sashi product is also with a confidence value of 0.959.
7. If you buy Lalita products, then the possibility of buying Sashi products is also with a confidence value of 0.961.
8. If you buy Alice and Cassandra products, then the possibility of buying Sashi products is also with a confidence value of 0.974.
9. If you buy Acio products, then the possibility of buying Sashi products is also with a confidence value of 0.987.

Conclusion

Based on the research that has been carried out, a conclusion can be drawn from this study, namely that the data of clothing sales transactions at the Siny.co Store can be carried out association rules using the FP-Growth algorithm in the RapidMiner application. From the application of the FP-Growth algorithm model, the sales transaction data of the Siny.co Shop clothing was generated by 9 rules, including products that can be recommendations for the Siny.Co Shop, namely Sashi, Canna, Alice, Tartan, Aruna, Cassandra, Nala, Lalita, and Acio. Where these products can help the Siny.co Store to recommend products that are of interest to customers so that they can launch a promotional strategy for the Siny.co Store in the future.

Bibliography

Allam, Zaheer, & Dhunny, Zaynah A. (2019). On big data, artificial intelligence and smart cities. *Cities*, *89*, 80–91.

Brous, Paul, Janssen, Marijn, & Herder, Paulien. (2020). The dual effects of the Internet of Things (IoT): A systematic review of the benefits and risks of IoT adoption by organizations. *International Journal of Information Management*, *51*, 101952.

Hasan, Firman Noor, Aziz, Achmad Sufyan, & Nofendri, Yos. (2023). Utilization of Data Mining on MSMEs using FP-Growth Algorithm for Menu Recommendations. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *22*(2), 261–270.

Hilman, Kirana Anastasya Afika Putri Hilman. (2022). Analisa Data Penjulan pada Toko Kelontong Musyawarah Menggunakan Algoritma Apriori. *Journal of Applied Computer Science and Technology*, *3*(2), 221–227.

Kana, Fiqal, Ramadhan, Mukhlis, & Mahyuni, Rina. (2022). Implementasi Data Mining Menganalisa Pola Penjualan Rempah-Rempah Menggunakan Metode Fp-Growth. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(4), 557–564.

Kumar, Mrinabh, & Dubey, Animesh Kumar. (2023). An analysis and literature review of algorithms for frequent itemset mining. *International Journal of Advanced Computer Research*, *13*(62), 1.

Merliani, Nanda Nurisya, Khoerida, Nur Isnaeni, Widiawati, Neta Tri, Triana, Latifah Adi, & Subarkah, Pungkas. (2022). Penerapan Algoritma Apriori Pada Transaksi Penjualan Untuk Rekomendasi Menu Makanan Dan Minuman. *J. Nas. Teknol. Dan Sist. Inf*, *8*(1), 9–16.

Salu, Nerwan, Michael, Aryo, Padang, Samuel Yakobus, & Adda, Muhammad Sofwan. (2022). Prediksi Persediaan Sepeda Motor Pada Dealer Yamaha Jaya Baru Motor Mengunakan Metode Frequent Pattern (FF-Growth). *INFINITY: UKI Toraja Journal of Information Technology*, *2*(1).

Sari, Kumala, Muhammad Syahril, S. E., Kom, M., & Suharsil, S. E. (2021). Penrapan Data Mining Untuk Pengelompokkan Data Member Restoran Dengan Menggunkan Metode Hierarki Agglomerative Single Lingkage Pada Restoran Al-Jazeerah Medan. *Jurnal Cyber Tech*, *4*(2).

Saura, Jose Ramon. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92–102.

Shawkat, Mai, Badawi, Mahmoud, El-ghamrawy, Sally, Arnous, Reham, & El-desoky, Ali. (2022). An optimized FP-growth algorithm for discovery of association rules.

The Journal of Supercomputing, *78*(4), 5479–5506.

Wardani, Ni Wayan. (2020). *Penerapan data mining dalam analytic CRM*. Yayasan Kita Menulis.

Wu, Yilan, & Zhang, Jing. (2023). *Retraction Note: Building the electronic evidence analysis model based on association rule mining and FP-growth algorithm*. Springer.