Marketing Strategy Using Frequent Pattern Growth

Nazori Suhandi[1] , Rendra Gustriansyah[2]*
[1][2]Universitas Indo Global Mandiri, Indonesia
[1]nazori@uigm.ac.id, [2]rendra@uigmm.ac.id

ABSTRACT

The biggest problem faced by printing companies during the Covid-19 pandemic was that the number of orders was unstable and tends to decrease, which had the potential to harm the company. Therefore, various appropriate marketing strategies were needed so that the number of product orders was relatively stable and even increases. The impact was that the company could survive and continued to grow. This study aimed to assist company managers in developing appropriate marketing strategies based on association rules generated from one of the data mining methods, namely the Frequent Pattern Growth (FP-Growth) method. The case study of this research was a printing company where there was no similar research that used a printing company's dataset. This study produced nine association rules that meet a minimum of 25% support and a minimum of 60% confidence, but only two association rules that had a high positive correlation, namely for a custom paper bag and banner products. Therefore, several marketing strategies were suggested that could be used as guidelines for companies in managing sales packages and giving special discounts on a product. The results of this study are expected to trigger an increase in the number of product orders because this study tried to find the right product for consumers and did not try to find the right consumers for a product.

Keywords: Association Rules; Data Mining; FP-Growth; Marketing Strategy; Printing

INTRODUCTION

During the current Covid-19 pandemic, the biggest problem faced by printing companies is the number of orders that are unstable and tend to decrease, thereby reducing the company's turnover. This can be detrimental to the company if management cannot anticipate it. Company management is required to be able to make the right marketing strategy so that the number of product orders is relatively stable so that the company can survive and can continue to grow. One way that can be applied is knowing consumer desires and providing product choices that suit consumer needs so that it can have an impact on increasing the number of orders.

The consumer wants and needs can be identified by analyzing the pattern of consumer ordering transactions. Several data mining methods can be used to analyze patterns from consumer order transaction datasets, namely Apriori (Lubis & Hasugian, 2020), FP-Growth (Han, Pei, & Yin, 2000), Eclat (Zaki, 2000), GSP (Rendra Gustriansyah, Sensuse, & Ramadhan, 2015; Srikant & Agrawal, 1996), and others. Meanwhile, this study will use the Frequent Pattern Growth (FP-Growth) method, because the FP-Growth method can analyze order transaction patterns faster than the Apriori algorithm (Abdullah, 2018; Hossain, Sattar, & Paul, 2019) and Generalized Sequential Pattern (GSP) (Destrilia, Primartha, Sukemi, & Wijaya, 2020), and produced many association rules compared to Eclat (Siregar, Kusuma, Kuncoro, & Suliswaningsih, 2018). FP-Growth can extract correlations between products in the dataset so that it can identify various association rules between products (items).

Therefore, this study aims to obtain the rules of association between products by applying the FP-Growth method as a guide in developing marketing strategies for printing companies. The association rules are the result of analysis of consumer ordering transaction patterns. The results of the analysis can be in the form of products that are often ordered simultaneously, the level of certainty of a product being ordered when other products are ordered, and products that are less attractive to consumers. The marketing strategy that can be applied is in the form of determining product packages, product discounts, or complementary products (Abdullah, 2018).

The results of this study are expected to provide guidelines for company management in managing the right product sales package and giving discounts on a product so that it can trigger an increase in the number of orders.

* Corresponding author

LITERATURE REVIEW

Based on the meta-files of Scopus and Google Scholar in the last three years, there are 372 unique papers (not duplicates) that use the FP-Growth method and 3 unique papers (Ariestya, Supriyatin, & Astuti, 2019; Destrilia et al., 2020; Putro & Gunawan, 2019) who use the FP-Growth method as a marketing strategy, but no related research has been found that is applied to printing companies.

Destrilia's research uses a gift retail dataset (Destrilia et al., 2020), Putro uses a hydroponic retail dataset (Putro & Gunawan, 2019), and Arestya using a staple retail dataset (Ariestya et al., 2019). Meanwhile, this research will use a retail printing dataset.

Frequent Pattern Growth

Frequent Pattern Growth (FP-Growth) is a data mining method introduced by Han et al. (2000) to extract association rules between products (items) in a transactional dataset (Han et al., 2000). FP-Growth uses the Frequent Pattern Tree (FP-tree) to generate the frequency of occurrence (frequent itemset) of each product (Novita, Mustakim, & Salisah, 2021).

FP-tree is a tree structure where each branch contains nodes that store item information along with its frequency and each leaf is the item with the smallest frequency. Before building the FP-tree, the items in each transaction must be sorted according to their frequency of occurrence first so that the item search process is faster. All items that do not meet the specified minimum support (min_supp) can be eliminated.

The FP-tree that is built has a header table that contains information about items, frequencies, and information for each item in the tree. The list of items in the header table is sorted in descending order by frequency, starting from the item with the highest frequency to the lowest. Furthermore, the process of finding the frequency of the itemset is only done through this FP-tree.

FP-Growth uses two important parameters to generate association rules, namely support and confidence, and uses a lift ratio to evaluate the strength (validity) of the association rules that have been generated.

Support is an indicator of the frequency of occurrence of an item from the transaction dataset. The support value for an item, for example, item A, can be obtained using (1) (Meida, Rini, & Sukemi, 2019).

$$supp(A) = \frac{\Sigma \ transaction \ containing \ A}{\Sigma \ transaction} \qquad (1)$$

Meanwhile, to find the support value of the 2-itemset, for example, items A and B, it can be calculated using (2) (Novita et al., 2021).

$$supp(A \cup B) = \frac{\Sigma \ transactions \ containing \ A \ and \ B}{\Sigma t \ ransactions} \qquad (2)$$

Confidence is an indicator that shows the association rules between two itemsets conditionally where the probability of an item is certain to be purchased together with the purchase of several other items. The confidence value of a rule, for example, A B, in a transaction can be calculated using (3) (Meida et al., 2019).

$$conf(A \Rightarrow B) = \frac{supp(A \cup B)}{supp(A)} = \frac{\Sigma transactions \ containing \ A \ and \ B}{\Sigma transactions \ containing \ A} \qquad (3)$$

The support and confidence values are in the interval (0, 1) (Destrilia et al., 2020).

Lift ratio is a parameter used to measure the level of strength (validity) of the association rule that has been generated (Destrilia et al., 2020). The lift ratio value is in the interval (0, +∞). The higher the lift ratio value, the stronger the association rules produced (Destrilia et al., 2020). The lift ratio value > 1 is a good lift ratio value because it shows that items A and B are positively correlated or item B is positively dependent on item A. The lift ratio can be calculated using (4) (Destrilia et al., 2020; Meida et al., 2019)

* Corresponding author

$$Lift\ ratio\ (A \rightarrow B) = \frac{supp(A \cup B)}{supp(A) * supp(B)} = \frac{conf(A \Rightarrow B)}{supp(B)} \qquad (4)$$

The FP-Growth steps to obtain detailed association rules are as follows (Han et al., 2000):
Step 1. Scan the transactional dataset to extract the frequency of each item (frequent itemset).
Step 2. Sorting the itemset in descending order based on the support (frequency) of each item.
Step 3. Determine the minimum support and confidence. If the support value of the item < minimum support (min_supp) then the item will be eliminated.
Step 4. Arrange the transaction dataset based on the priority of the itemset.
Step 5. Build the FP tree.
Step 6. Determine the Conditional Pattern Base, Conditional FP-tree, and Frequent itemset.
Step 7. Calculate the lift ratio for each rule.
Step 8. Generate valuable association rules.

METHOD

Research Framework

The stages of this research can be seen in Fig. 1 where the sample dataset is obtained from a printing company with a minimum support value of 25% and a minimum confidence value of 60%. In addition, this research will be designed using a UML (R. Gustriansyah, Suhandi, & Antony, 2019; Setiawan, Sumitro, & Gustriansyah, 2019).
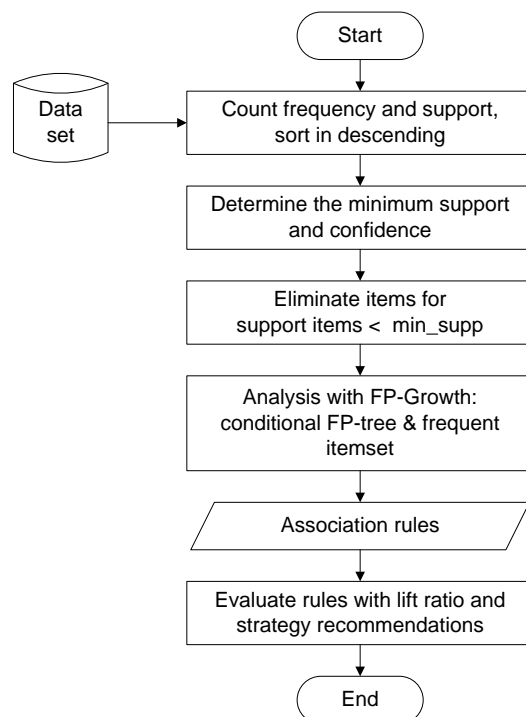

Fig. 1 Research Stages

Dataset

This study uses a transactional dataset from printing companies as shown in Table 1. The first column represents the transaction ID and the second column represents the items (products) purchased together by consumers in one transaction.

* Corresponding author

Table 1
Order Transactional Dataset

Transaction ID	Product Name
T21-0001	thank you card, paper bag custom
T21-0002	paper bag custom, banner, plakat, sertifikat, amplop custom
T21-0003	thank you card, banner, plakat, souvenir, undangan
T21-0004	thank you card, plakat, souvenir
T21-0005	paper bag custom, thank you card, banner, desain grafis
T21-0006	banner, plakat, thank you card, paper bag custom
T21-0007	thank you card, plakat
T21-0008	paper bag custom, thank you card, banner
T21-0009	thank you card, paper bag custom, plakat
T21-0010	paper bag custom, banner, souvenir

The support or frequency of occurrence of each product (item) can be seen in Table 2. All items are sorted in descending order based on their frequency/support. Items that have a min_supp < 25% or a frequency of < 2 (i6 to i10 items) will be eliminated.

Table 2
Frequency of Ordering Each Product

Item	Product Name	Frequency	Support
i1	thank you card	8	80%
i2	paper bag custom	7	70%
i3	banner	6	60%
i4	plakat	5	50%
i5	souvenir	3	30%
i6	invitation card	1	10%
i7	placard	1	10%
i8	graphic design	1	10%
i9	certificate	1	10%
i10	custom envelope	1	10%

Then, all items from Table 2 are arranged based on the priority of occurrence frequency for each order transaction as shown in Table 3.

Table 3
Transaction Dataset by Priority

Transaction ID	Itemset
T21-0001	{i1,i2}
T21-0002	{i2,i3,i4}
T21-0003	{i1,i3,i4,i5}
T21-0004	{i1,i4,i5}
T21-0005	{i1,i2,i3}
T21-0006	{i1,i2,i3,i4}

* Corresponding author

Transaction ID	Itemset
T21-0007	{i1}
T21-0008	{i1,i2,i3}
T21-0009	{i1,i2,i4}
T21-0010	{i2,i3,i5}

Based on the itemset from Table 3, an FP-tree and table header was built as can be seen in Fig. 2. The validation of the FP-tree that has been formed can be done by calculating the number of item frequencies that must be equal to the number of item frequencies in Table 3. In this study, the number of item frequencies from both is the same, which is 29 items.
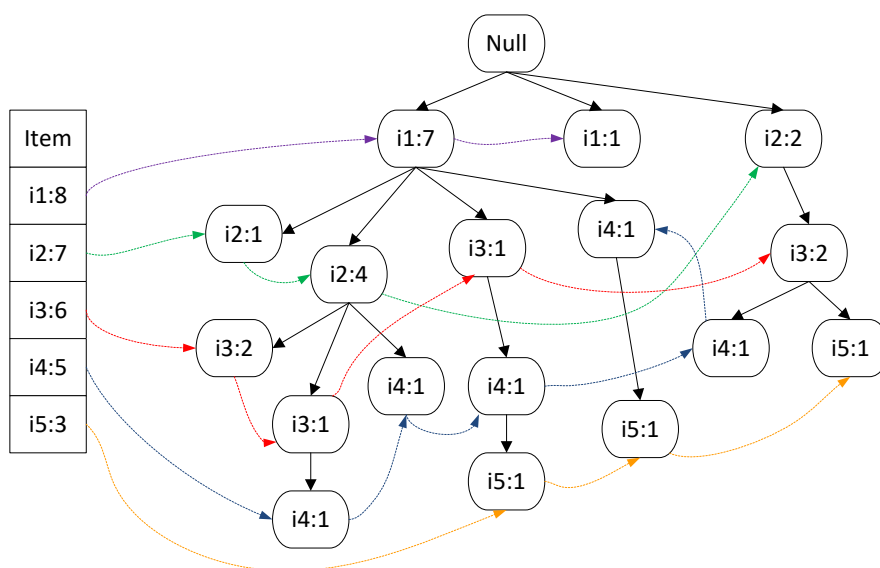

Fig. 2 FP-tree (Min_supp ≥ 25%)

Table 4 is a conditional FP-tree and frequent itemset obtained from Fig. 2 with a min_supp ≥ 25%. Nodes with min_supp < 25% will be eliminated from the conditional FP-tree so that the appropriate pattern or frequent itemset and its support can be generated.

Table 4
Frequent Itemset from FP-Growth (Min_supp ≥ 25%)

Item	Conditional *FP-tree*	*Frequent itemset*
i5	<i1:2, i4:2>	{i1,i5:2}, {i4,i5: 2}, {i1,i4,i5: 2}
i4	<i1:4, i2:2>	{i1,i4:4}, {i2,i4:3}, {i3,i4:2}, {i2,i3,i4:2}
i3	<i1:4, i2:3>, <i2:2>	{i1,i3:4}, {i2,i3:5}, {i1,i2,i3:3}
i2	<i1:5>	{i1,i2: 5}

RESULT

Based on the frequent itemset from Table 4, all combinations of association rules for the frequent 2-itemset are compiled. Then, to evaluate the combination of association rules that have been generated, lift ratio - (4) is used with min_supp value ≥ 25% and confidence ≥ 60%.

* Corresponding author

The lift ratio value for each association rule will be sorted in descending order based on the confidence value as shown in Table 5. All combinations of association rules that have a min_supp value < 25% and confidence < 60% will be ignored (association rules are colored in gray) so that we get the nine rules of association that meet.

Table 5
Frequent 2-Itemset

Rules	*Support*	*Confidence*	*Lift ratio*
i3 ⟹ i2	**50%**	**83%**	**1.19**
i4 ⟹ i1	40%	80%	1
i2 ⟹ i3	**50%**	**71%**	**1.19**
i2 ⟹ i1	50%	71%	0.89
i3 ⟹ i1	40%	67%	0.83
i5 ⟹ i1	20%	67%	0.83
i5 ⟹ i4	20%	67%	1.33
i1 ⟹ i2	50%	63%	0.89
i4 ⟹ i2	30%	60%	0.86
i1 ⟹ i4	40%	50%	1
i1 ⟹ i3	40%	50%	0.83
i2 ⟹ i4	30%	43%	0.86
i4 ⟹ i5	20%	40%	1.33
i4 ⟹ i3	20%	40%	0.67
i3 ⟹ i4	20%	33%	0.67
i1 ⟹ i5	20%	25%	0.83

DISCUSSIONS

Table 5 shows that the combination of association rules that have a high positive correlation only consists of two rules, namely i3 ⟹ i2 and i2 ⟹ i3 where the lift ratio value is > 1. This means that custom paper bag (i2) and banner (i3) are a combination of a good products. One of the marketing strategies that can be applied to trigger an increase in the number of orders for these two products is to make these two products in a sales offer package accompanied by a discount.

The results of this study cannot be compared directly with similar studies because of the case studies and the use of different datasets. For example, Destrilia's research (Destrilia et al., 2020) produces more rules, namely 8 rules but using a min_supp value of only 4%. Likewise with Putra's research (Putro & Gunawan, 2019) produces 21 rules but with a min_supp value of only 5%. This is because the determination of the min_supp value is affected by the size of the dataset and will also have an impact on the number of rules generated. The higher the min_supp and confidence values used, the fewer rules are formed.

Meanwhile, the marketing strategy that can be applied to a combination of association rules that has a confidence value of 60% but a lift ratio value of < 1 is to offer a second item (product) as a recommended product to order when consumers order the first item (product). For example, when a consumer buys a placard (i4), the system offers/recommends a thank you card (i1) and a custom paper bag (i2). Then, to facilitate the promotion, the second product can be displayed adjacent to the first product. In addition, the first and second products can also be used as a sales offer package accompanied by discounts. For items (products) that have a support value of < 25%, they must be actively promoted and given bigger special discounts as an attraction for consumers.

CONCLUSION

Analyzing the pattern or frequent itemset of a transactional dataset is the main goal of the FP-Growth method. This study resulted in nine association rules that meet 25% min_supp and greater than 60% confidence, but only two

* Corresponding author

association rules that have a high positive correlation are custom paper bags (i2) and banners (i3). Based on the results of testing and discussion, it can be concluded that the combination of association rules generated by the FP-Growth method, which is the result of the analysis of product ordering transaction patterns, can be used as a guide in developing the right marketing strategy for printing companies. The implementation of this marketing strategy is expected to trigger an increase in the number of product orders.

Further research can use various other association rule mining methods such as Apriori Hybrid, Fast Eclat, Fold Growth, and so on to get faster and more representative calculation results.

REFERENCES

Abdullah, A. (2018). Rekomendasi Paket Produk Guna Meningkatkan Penjualan dengan Metode FP-Growth. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, *4*(1), 21. https://doi.org/10.23917/khif.v4i1.5794

Ariestya, W. W., Supriyatin, W., & Astuti, I. (2019). Marketing Strategy for The Determination of Staple Consumer Products Using FP-Growth and Apriori Algorithm. *Jurnal Ilmiah Ekonomi Bisnis*, *24*(3), 225–235. https://doi.org/10.35760/eb.2019.v24i3.2229

Destrilia, Primartha, R., Sukemi, & Wijaya, A. (2020). Online Retail Marketing Recommendation System Based on Generalized Sequential Pattern Algorithm and FP-Growth Algorithm. In *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*. Paris, France: Atlantis Press. https://doi.org/10.2991/aisr.k.200424.053

Gustriansyah, R., Suhandi, N., & Antony, F. (2019). The Design of UML-Based Sales Forecasting Application. *International Journal of Recent Technology and Engineering*, *7*(6), 1507–1511.

Gustriansyah, Rendra, Sensuse, D. I., & Ramadhan, A. (2015). Decision support system for inventory management in pharmacy using fuzzy analytic hierarchy process and sequential pattern analysis approach. In *2015 3rd International Conference on New Media (CONMEDIA)* (pp. 1–6). IEEE. https://doi.org/10.1109/CONMEDIA.2015.7449153

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00* (pp. 1–12). New York, New York, USA: ACM Press. https://doi.org/10.1145/342009.335372

Hossain, M., Sattar, A. H. M. S., & Paul, M. K. (2019). Market Basket Analysis Using Apriori and FP Growth Algorithm. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCIT48885.2019.9038197

Lubis, A. E., & Hasugian, P. M. (2020). Implementation of Data Mining on Suzuki Motorcycle Sales in Gemilang Motor Prosperous with Apriori Algorithm Method. *Journal of Computer Networks, Architecture and High Performance Computing*, *2*(1), 23–29. https://doi.org/10.47709/cnapc.v2i1.353

Meida, A., Rini, D. P., & Sukemi, S. (2019). Pattern of E-marketplace Customer Shopping Behavior using Tabu Search and FP-Growth Algorithm. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *7*(4). https://doi.org/10.11591/ijeei.v7i4.1362

Novita, R., Mustakim, & Salisah, F. N. (2021). Determination of the relationship pattern of association topic on Al-Qur'an using FP-Growth Algorithms. *IOP Conference Series: Materials Science and Engineering*, *1088*(1), 012020. https://doi.org/10.1088/1757-899X/1088/1/012020

Putro, A. N. S., & Gunawan, R. I. (2019). Implementasi Algoritma FP-Growth untuk Strategi Pemasaran Ritel Hidroponik (Studi Kasus : PT. HAB). *Jurnal Buana Informatika*, *10*(1), 11. https://doi.org/10.24002/jbi.v10i1.1746

Setiawan, H., Sumitro, A. A. A., & Gustriansyah, R. (2019). The change data capture and the web application messaging protocol on the real time dashboard. *International Journal of Engineering and Advanced Technology*, *8*(4).

Siregar, A. K., Kusuma, B. A., Kuncoro, A. P., & Suliswaningsih. (2018). Perbandingan Algoritme FP-Growth dan Eclat untuk Analisis Pola Pembelian Konsumen pada Toko "X." In *Conference on Information Technology, Information System and Electrical Engineering (CITISEE)* (pp. 125–128). Purwokerto: Universitas AMIKOM Purwokerto.

* Corresponding author

Srikant, R., & Agrawal, R. (1996). Mining Sequential Patterns: Generalizations and Performance Improvements. In *5th International Conference on Extending Database Technology Avignon* (pp. 1–17). France.

Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, *12*(3), 372–390. https://doi.org/10.1109/69.846291

* Corresponding author