International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Implementation Data Mining Using FP-Growth Algorithm to Determine Customer Purchase Patterns

Ida Farida
Mercu Buana University, South Meruya No.1, Indonesia
Email address: dae.farida@mercubuana.ac.id

Abstract*—In building a business, a distributor company really needs the use of information technology, in order to support the smooth sales of food products provided. The problem that occurs is how to provide the best service to the market or customers to use the right business strategy to avoid losses to the company. Businesses must discover a method for boosting sales of the goods they offer. One way that can be done to determine a sales strategy is to use data mining techniques. The technique used in this case is the Fp-Growth algorithm, which is an algorithm that produces frequent itemsets which will later be used in the process of determining rules that can produce a choice. The Fp-Growth algorithm is a development of the Apriori algorithm. The Fp-Growth algorithm uses the concept of tree development in searching for types of goods that are frequently purchased (frequent itemsets).*

Keywords*— Sales, Data mining, Association rules, Frequent itemset, Fp- Growth.*

I. INTRODUCTION

The competition in the business world today is very tight, especially in the distributor business. E-commerce companies in building their businesses do not escape from using information technology to support the smooth sales of the products provided. It is undeniable that the use of information technology is a must at this time, considering the rapid and tight business competition in this field.

The problem that occurs is in terms of providing the best service to the market or company customers to use the right business strategy to avoid company losses. However, sometimes the company management is constrained in determining the strategy. Many factors cause this. One of the causal factors is the difficulty of producing analysis related to existing product sales data patterns, thus giving a negative impact on the results of product sales that go up and down.

One of the problems that occurs is that the most sold and interrelated products are still not well organized, so that consumer purchasing patterns for a selected product are not analyzed properly to determine the product stocking strategy to be marketed. Of course this causes the stocking strategy of goods to be unplanned in the future.

One of the techniques that can be used to determine consumer purchasing pattern strategies is data mining techniques. Data mining is an automatic process to find new information (knowledge) that has potential from a set of data (Kurniasih et al., 2022). Data mining techniques require algorithms to make calculations more complex, one of which is the FP-Growth algorithm. The FP-Growth algorithm is a development of the Apriori algorithm. So that the shortcomings of the Apriori algorithm are corrected by the FP-Growth algorithm.

Frequent Pattern Growth (FP-Growth) is one of the alternative algorithms that can be used to determine the most frequently occurring data set (frequent itemset) in a data set (Fahrin & Maulana, 2018). Research conducted by Herasmus using the FP-Growth algorithm in analyzing the Customer Service System, concluded that the support value given was an average of 80 percent and the confidence value was above 80 percent (Herasmus, 2017). So based on the description of the problem and previous research, this study will apply the FP-Growth algorithm in carrying out the data mining process to determine the consumer purchasing patterns of PT. Cipta Niaga Semesta (Mayora Group). From the resulting pattern, information will be produced that can be used by PT. Cipta Niaga Semesta (Mayora Group) to improve sales quality, service quality and profits by minimizing losses.

Based on the description of the background above, a problem formulation can be taken, namely how to apply the FP-Growth algorithm to determine customer purchasing patterns.

II. LITERATURE REVIEW

2.1. Definition of Association Rule

Association rules is a method that aims to find patterns that often appear among many transactions, where each transaction consists of several items so that this method will support the recommendation system by finding patterns between items in the transactions that occur. (Fadlina, 2024)

2.2. Definition the FP-Growth Algorithm

The FP-Growth algorithm is a development of the Apriori algorithm. The Frequent Pattern Growth (FP-Growth) algorithm is an alternative algorithm that can be used to determine the most frequently occurring data set (frequent itemset) in a data set. FP-Growth can find the frequency of itemsets with only a little access to the original database, and its approach is the most efficient. (Wiyana, 2018)

The FP-Growth algorithm can also avoid problems if the number of candidate itemsets is too large. The notion of tree construction is used by the FP-Growth algorithm to look for common itemsets. The FP-Growth algorithm is faster than the Apriori algorithm for this reason. The FP-Growth algorithm can directly extract frequent itemsets from the FP-Tree by using FP-Tree. Mining frequent itemsets using the FP-Growth algorithm will be done by generating a tree data structure. (Erwin, 2009)

III. Research Methodology

The research method used in this study is a qualitative approach. Qualitative methods are called new methods, because their popularity has not been long, called post-positivism methods. Qualitative research methods are research methods based on the philosophy of post-positivism, used to research natural objects (as initially as experiments) Researchers are crucial tools, data collection methods are based on triangulation (combination), inductive/qualitative analysis, and qualitative research findings place more emphasis on food than generalization (Sugiyono, 2014; 7). This research is qualitative by analyzing existing customer purchasing patterns by utilizing data mining methods and FP Growth algorithm calculations.

IV. Results and Discussion

The FP-Growth algorithm is distinguished by its use of a tree-like data structure known as the FP-Tree. The FP-growth method can directly extract frequent itemsets from the FP-Tree by using FP-Tree. Mining frequent itemsets using the FP-Growth algorithm will be done by generating a tree data structure or called FP-Tree. FP-Growth can be divided into 3 main stages, namely:

1. Conditional pattern base generation stage
2. Conditional FP-Tree generation stage, and
3. Frequent itemset search stage.

These three stages are the steps that will be taken to obtain frequent itemsets. The following are the manual calculation steps of the Fp-Growth algorithm. In the process of implementing the Fp-Growth algorithm, there are several stages that must be carried out. The table below is a manual calculation to find frequent itemsets containing 20 sample transaction data.

TABLE 1. To Mark Each Itemset

ITEM
(A) 2B pencil
(B) Black Pen
(C) Blue Pen
(D) Red Pen
(E) Eraser
(F) Pencil Sharpener
(G) 38 Sheet Notebook
(H) 58 Sheet Notebook
(I) A5 Picture Book
(J) A4 Drawing Book
(K) 15 cm ruler
(L) Ruler 30 cm
(M) Black Marker
(N) Red Marker
(O) Blue Marker
(P) Yellow Highlighter
(Q) Green Highlighter
(R) Plastic Map
(S) Small Sticky Notes
(T) Big Sticky Notes

TABLE 2. Transaction Data

TID	Item
1	{ A,B,D,F,G,H,I,K,N,P,R,S }
2	{ A,C,D,E,F,I,M,O,Q,R }
3	{ A,C,D,E,F,G,J,K,M,N,O,Q,S }
4	{ A,B,C,D,E,F,H,I,K,N,P,Q,S }
5	{ A,B,C,E,H,I,J,M,N,P,R }
6	{ B,C,G,H,I,K,M,O,P,Q,R }
7	{ C,D,J,M,O,P,R }
8	{ A,B,D,G,H,I,L,P,S }
9	{ A,C,D,E,,G,J,Q,R,S }
10	{ B,D,F,H,I,K,M,N,O,S,T }
11	{ A,B,C,E,N,Q,S }
12	{ A,B,C,G,H,I,L,O,Q,T }
13	{ A,B,C,E,G,K,L,R,T }
14	{ B,D,F,J,M,N,O,Q }
15	{ A,B,C,D,E,G,L,Q,T }
16	{ E,M,N,O,S }
17	{ A,D,E,F,H,I,M }
18	{ C,D,L,P,T }
19	{ C,E,L,N,P,Q }
20	{ B,D,G,J,M,N,T }
21	{ D,H,I,L,P,R,S }
22	{ A,C,G,K,N,R }
23	{ A,B,C,H,L,N,Q,R,S }
24	{ A,B,C,D,K,M }
25	{ A,C,D,E,H,J,P }
26	{ A,B,C,D,G,J,P,T }
27	{ A,B,D,F,P,T }
28	{ A,B,P,S }
29	{ A,C,D,F,G,N }
30	{ A,B,H,I }

After the sample data is prepared, the data will be converted into binomial format which can be seen in the table below:

TABLE 3. Sales Transactions in Binomial Format

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	1	0	1	0	1	1	1	1	0	0	1	0	1	0	1	1	1	1	1
1	0	1	1	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0
1	0	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1	0	1	0
1	1	1	1	1	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0
1	1	1	0	1	0	0	1	1	1	0	0	1	1	0	1	0	1	0	0
0	1	1	0	0	0	1	1	1	0	1	0	1	0	1	1	1	1	0	0
0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	1	0	1	0	0
1	1	0	1	0	0	1	1	1	0	0	1	0	0	0	1	0	0	1	0
1	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	1	1	1	0
0	1	0	1	0	1	0	1	1	0	1	0	1	1	1	0	0	0	1	1
1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0
1	1	1	0	0	0	1	1	1	0	0	0	1	0	1	0	1	0	0	1
1	1	1	0	1	0	1	0	0	0	1	1	0	0	0	0	0	1	0	1

The itemset data in the table above explains the product items that are used as attributes to be executed, each product sold in each transaction will be marked with the number 1 or have a true value, while products that are not sold will be marked with the number 0 or have a false value.

After the frequency of each item appears, the products are sorted based on the largest to smallest frequency, and can be seen in the table below:

TABLE 4. Frequency Table and Support Values

Product	Frequency	Support Value
2B Pencil(A)	21	105 %
Blue Pen(C)	20	100%
Red Pen (D)	19	95%
Black Pen(B)	18	90%
Red Marker (N)	13	65%
Yellow Stabilo (P)	13	65%
Eraser(E)	12	60%
38 Sheets Notebook (G)	12	60%
58 Sheets Notebook (H)	12	60%

A5 Picture Book (I)	11	55 %
Black Marker (M)	11	55 %
Small Sticky Notes (S)	11	55 %
Green Highlighter (Q)	10	50%
Plastic Map(R)	10	50%
Big Sticky Notes (T)	8	40%
Blue Marker (O)	8	40%
Ruler 30 cm (L)	8	40%
Pencil Sharpener (F)	7	35 %
A4 Drawing Book (J)	7	35 %
Ruler 15 cm (K)	7	35 %

After calculating the frequency of occurrence of each item, it is known that the items above the support count value = 55% or items with a frequency above 11 are A, C, D, B, N, P, E, G, H, I, M, and S. These 12 items will have an influence and will be included in the fp-tree, the rest (Q, R, T, O, L, F, J, and K) are not used because they do not have a significant influence as shown in Table 5.

TABLE 5. Transactions adjusted to frequent list

TID	Item
1	{ A,B,D,G,H,I,N,P,S }
2	{ A,C,D,E,I,M }
3	{ A,C,D,E,G,M,N,S }
4	{ A,B,C,D,E,H,I,N,P,S }
5	{ A,B,C,E,H,I,M,N,P }
6	{ B,C,G,H,I,M,P }
7	{ C,D,M,P }
8	{ A,B,D,G,H,I,P,S }
9	{ A,C,D,E,G,S }
10	{ B,D,H,I,M,N,S }
11	{ A,B,C,E,N,S }
12	{ A,B,C,G,H,I }
13	{ A,B,C,E,G }
14	{ B,D,M,N }
15	{ A,B,C,D,E,G }
16	{ E,M,N,S }
17	{ A,D,E,H,I,M }
18	{ C,D,P }
19	{ C,E,N,P }
20	{ B,D,G,M,N }
21	{ D,H,I,P,S }
22	{ A,C,G,N }
23	{ A,B,C,H,N,S }
24	{ A,B,C,D,M }
25	{ A,C,D,E,H,P }
26	{ A,B,C,D,G,P }
27	{ A,B,D,P }
28	{ A,B,P,S }
29	{ A,C,D,G,N }
30	{ A,B,H,I }

Then the next step is to form a tree which is the next step in the fp-growth algorithm process based on table 5. The creation of an fp-tree that starts from TID 1 is {A,B,D,G,H,I,N,P,S}.

TABLE 6. Formed Rules

Rules	Suport	Confidence
A,B	143%	205%
AIR CONDITIONING	143%	205%
A,D	123%	176%
B,D	117%	167%
B,H	117%	167%
C,E	97%	138%
H,I	97%	138%
B,C	90%	129%
CD	90%	129%
D,H	90%	129%
A,E	87%	124%
B,I	80%	114%
D,G	77%	110%
AH	73%	105%
D,E	73%	105%
IN	73%	105%
C,G	67%	95%
C,H	67%	95%
A,G	60%	86%
B,N	50%	71%
G,H	50%	71%
C,N	47%	67%
EH	47%	67%
A,I	43%	62%
B,M	43%	62%
C,I	43%	62%
D,M	43%	62%
D,N	43%	62%
D,P	43%	62%
C,M	40%	57%
E,N	40%	57%
H,N	40%	57%
B,P	37%	52%
E,I	37%	52%
E,M	37%	52%
G,I	37%	52%
MOBILE PHONE	37%	52%
I,M	33%	48%
I,N	33%	48%
I,P	33%	48%
M N	33%	48%
A,N	30%	43%
H,M	30%	43%
US	27%	38%
C,P	27%	38%
M,S	27%	38%
B,S	23%	33%
D,S	23%	33%
E,G	23%	33%
G,M	23□	33%
G,N	20%	29%
G,P	20%	29%
N,P	20%	29%
N,S	20%	29%
C,S	17%	24%
I,S	17%	24%
P,S	17%	24%
A,P	13%	19%
E,P	135	19%
ICE	13%	19%
A,M	10%	14%
G,S	10%	14%
M,P	10%	14%

From the confidence calculation of the pattern formed above, the Association Rule that meets the confidence requirement ≥0.75 is: H → A = 0.75 (if the consumer buys a notebook, then buys a pencil), G → C = 0.75 (if the consumer buys a notebook, then buys a pen).

To prove the truth of the analysis results, a testing process is required to test the truth of the data processing results that have been done manually previously. For this testing process, we can use one of the application software such as Rapidminer, with the following steps.

The dataset consists of 30 data records stored in the Microsoft Excel application and will be tested using Rapidminer 9.3.1 software to see whether the search results for frequent itemsets are the same or not.

Figure 1 is the process of importing dataset into rapidminer.

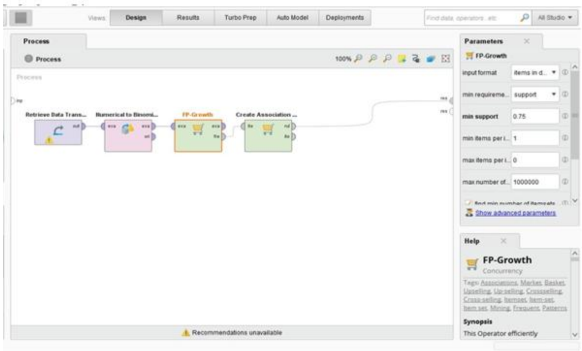

Figure 1. Support and Confidence Using Rapidminer from Sample Sales Data

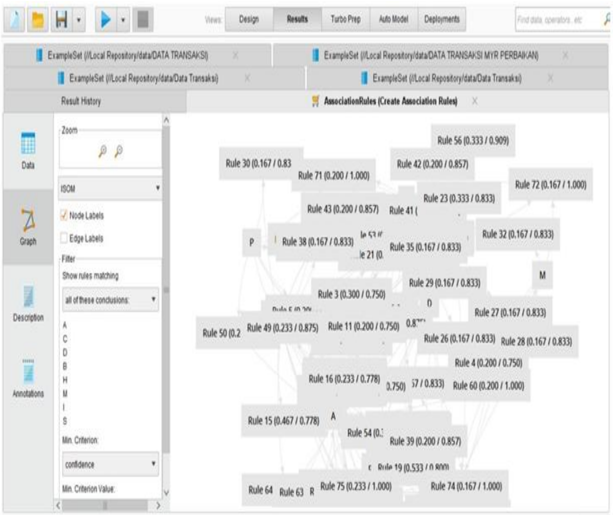

Figure 2. Sales Data Sample Rule Graphic Using Rapidminer

After explaining the rules above, it can be concluded that consumers tend to buy items that are related to each other, such as: if a consumer buys a book, they also buy a pen, with a confidence value of 0.75.

Then the final result obtained from the rule is a new knowledge about consumer purchasing patterns that have been rarely known. This can be used to help companies maintain their quality well for consumers, by means of products that must remain available when consumers need them. For that, companies must always provide products that consumers want.

V. Conclusion

Based on the data and discussion results, the following conclusions can be drawn:

1. Overall, from the sales sample data, 88 rules were obtained, consisting of 53 association rules that met the support threshold of 75% and 4 rules that meet 90% confidence.
2. New insights into consumer behavior are gained from the regulations acquired. Using this information, we may develop discount bundles for product purchase patterns with the greatest confidence value and make inventory judgments.
3. Companies may use the fp-growth method to keep an eye on their inventory of items that customers buy frequently, which will help prevent supply shortages.

Refferences

[1.] Nuraisana. (2019). Decision support system for determining recipients of the Hope Family Program (PKH) aid using the analytical hierarchy process method. *Jurnal Mantik Penusa, 3*(1), 129–137.

[2.] Parhusip, J. (2019). Application of analytical hierarchy process (AHP) method in designing decision support system for selection of candidates for recipients of non-cash food assistance (BPNT) in Palangka Raya City. *Jurnal Teknologi Informasi: Jurnal Ilmiah dan Aplikasi Bidang Teknologi Informasi, 13*(2), 18–29. https://doi.org/10.47111/jti.v13i2.251

[3.] Solihin, A. S., Gustian, D., Sembiring, F., Muslih, M., & Destria, N. A. (2021). Decision support system for cash social assistance recipients using technique for order preference by similarity to ideal solution method. *Jurnal Rekayasa Teknologi Nusa Putra, 8*(1), 1–9.

[4.] Martyani, E., & Santoso, S. (2019). Decision support system for choosing a hangout place using the analytical hierarchy process method. *Jurnal V-Tech (Vision Technology), 2*(1), 14–23. https://doi.org/10.35141/jvt.v2i1.455

[5.] Suprapto, B., & Sujoni, A. (2019). Decision support system for prospective recipients of Family Hope Program (PKH) using analytical hierarchy process (AHP) method. *Journal of Information and Computer, 7*(2), 47–56.

[6.] Fauzi, M. A. (2016). *Decision support system*. [Manuscript].

[7.] Laisouw, A. R., Lutfi, S., & Tempola, F. (2019). Decision support system for providing Family Hope Program (PKH) assistance to the poor in Ternate City using the AHP method. *JIKO (Journal of Information and Computer), 2*(1), 34–60. https://doi.org/10.33387/jiko.v2i1.973

[8.] Ministry of Social Affairs. (2021). Get to know the non-cash food assistance program. Kemensos.go.id. https://kemensos.go.id/uploads/topics/15664651387355.pdf

[9.] Serrano, N., Gallardo, G., & Hernantes, J. (2015). Infrastructure as a service and cloud technologies. *IEEE Software, 32*(2), 30–36.

[10.] Beer, M. I., & Hassan, M. F. (2017). Adaptive security architecture for protecting RESTful web services in enterprise computing environments. *Service Oriented Computing and Applications, 12*(2), 111–121. https://doi.org/10.1007/s11761-017-0221-1

[11.] Ardiansyah, H. I. (2017). *Development of REST API for Indoexplore.id tourism marketplace information system*.

[12.] Kusuma, I. (2017). How to quickly create API documentation with APIDOCjs. Retrieved from https://indrakusuma.web.id/2017/11/07/Cara-Mudah-Membuat-Dokumentasi-API-dengan-apidocjs

[13.] Perkasa, M. I., & Setiawan, E. B. (2018). Development of community data web service using REST API with access token. *Journal Name*, *X*(1), 19–26.

[14.] Pratama, Y. E. B. (2017). Implementation of REST web service for mobile-based reseller orders. *Journal Name*, *X*(Issue), 16–37.

[15.] Rulloh, et al. (2017). Implementation of REST API on Android-based flag raising guide application. *Teknikom, 1*(2), 85–89.

[16.] Sari, Y. S. (n.d.). Implementation of naïve Bayes classifier to determine disease.

[17.] Kaburuan, E. R., Sartika, Y., & Agustina, I. (2022). Sentiment analysis on product reviews from Shopee marketplace using the naïve Bayes classifier. *Journal Name*, *13*(1), 150–159.