Journal of Computer Science and Information Technology
https://jcsitech-upiyptk.org/ojs
2024 Volume 10 Issue 2 Pages 44-49 e-ISSN: 2502-1486

Application of the FP-Growth Algorithm in Consumer Purchasing Pattern Analysis

Indah Dwi Putri*, Yuhandri, Romi Hardianto
Putra Indonesia University YPTK Padang, Jl. Raya Lubuk Begalung Padang, West Sumatra – 25221, Indonesia
* indahdwiputri321@gmail.com

Abstract

Technology is currently used in various ways, one of which is businesses engaged in selling daily products. The right marketing strategy makes knowledge of consumer shopping patterns important to study because consumers are the main actors in carrying out transactions. The more diverse the types of goods sold in a company, the more diverse the resulting consumer spending patterns will be. Data mining is an analysis process that is carried out automatically on complex and large amounts of data to obtain patterns or trends that are generally not realized. The FP-Growth algorithm is an alternative algorithm that can be used to determine the data set that appears most frequently (frequent itemset) in a data set. The method used in this research is the FP-Growth method which is implemented in the PHP programming language and MySQL as the database. Designing a data mining program using the FP-Growth method can analyze and manage consumer purchasing patterns based on goods purchased simultaneously. The data processed in this research is transaction data that has been processed into information so as to gain knowledge in calculating stock of goods sourced from the owner of Toko Asra. From testing this method, results were obtained from the 10 transactions in December 2021, by limiting the minimum support value to 0.2 and minimum confidence to 0.75, 33 patterns of consumer shopping habits were obtained, meaning that 33 products were most frequently purchased by consumers. Designing a data mining program using the FP-Growth method can help analyze consumer purchasing patterns based on items purchased simultaneously. The results of frequent itemset calculations can help find a sequence of combinations that can be used as product recommendations in business decisions.

Keywords: Marketing, Stock, FP-Growth, Costumer, Purchasing Patterns

1. Introduction

Information technology is currently developing rapidly and this development cannot be separated from human development. Technology is currently used in various ways, one of which is businesses engaged in selling daily products [1]. Business competition in world trade through a free market economy followed by advances in information technology has brought companies to an increasingly fierce and open level of competition in meeting increasingly high consumer demands. So it forces the perpetrators to always think about strategies and breakthroughs on how to survive and develop their business [2].

Increasingly fierce competition requires large and small companies to have a system that can help develop business quality. Every company tries to implement new strategies in marketing their products by increasing what they earn by increasing the number of sales transactions that occur in their company continuously [3][4]. Appropriate marketing strategies make knowledge of consumer shopping patterns important to study because consumers are the main actors in carrying out transactions [5][6]. The more diverse the types of goods sold in a company, the more diverse the resulting consumer spending patterns will be. The Asra Shop is a shop that sells a variety of products that have the potential to observe consumer shopping patterns.

Buk Asra's Micro, Small and Medium Enterprises (MSMEs) is a basic food shop located on Jl. Raya Kelok Kuranji Padang. The problem that occurs at Toko Asra is that there is a shortage in one product and an excess in another product. The daily sales transaction process that occurs is not utilized properly so that information cannot be obtained that can be used as a strategy to increase sales.

Therefore, so that sales transaction data can be useful and useful, a data development method (data mining) is needed, especially in the use of the FP-Growth (Frequent Pattern Growth) algorithm, in utilizing sales transaction data obtained from goods purchased simultaneously by consumers at the time of purchase. shop. This data is used to find out the goods that consumers often buy to increase their stock to meet market demands and reduce stocking of goods that are not selling well on the market. The store also uses this data to arrange the placement of goods according to consumer behavior when shopping.

Accepted: 02-07-2024 | Revised: 03-11-2024 | Published: 04-30-2024 | doi: 10.35134/jcsitech.v10i2.99

Data mining is an analysis process that is carried out automatically on complex and large amounts of data to obtain patterns or trends that are generally not realized [7][8].

Market basket analysis (MBA) is a methodology for analyzing consumer shopping patterns by finding associations between several different items that consumers place in the shopping basket purchased in a particular transaction. The purpose of market basket analysis is to find out which products are likely to be purchased simultaneously [9][10].

One technique in data mining that is well known and suitable for MBAs is association rules. Association rule is a data mining technique that is useful for finding the most important or interesting correlation or pattern from a set [11].

Association rule is a form of pattern produced by Data Mining. Association Rule can be used to find relationships or cause and effect. Whether an association is important or not can be determined using two benchmarks, namely: support and confidence. Support (supporting value) is the percentage of the combination of items in the database, while confidence (certainty value) is the strength of the relationship between items in the association rules [12]. Association rules require repeated reading of transaction data with large amounts of transaction data to find different relationship patterns and are time consuming and have large computational costs. Therefore, an efficient algorithm is needed to find these patterns.

FP-Growth algorithm is an alternative algorithm that can be used to determine the data set that appears most frequently (*frequent itemset*) in a data set. FP-Growth uses the approach used in the a priori algorithm [13][14] . The FP-Growth algorithm uses the concept of tree development. What is usually called FP-Tree, in searching for frequent itemsets instead of using generated candidates as is done in the a priori algorithm [15][16]. By using this concept, the FP-Growth algorithm becomes faster and is divided into three main stages, namely:

1. *conditional pattern base* generation stage , *the conditional pattern base* is *a sub database* that contains *prefix paths* (prefix paths) and *suffi patterns* (suffix patterns). Generation of the *configurational pattern base* is obtained through *the FP-Tree* which has been built previously.
2. *Conditional FP-Tree* generation stage , at this stage , *the support count* of each item in each *conditional pattern base* is added up, then every *item* that has a *support count* greater than or equal to *the minimum support count* will be generated with *a conditional FP-Tree.*
3. *Frequent itemset* search stage , if the conditional *FP-Tree* is a single path , then *the frequent itemset is obtained* by combining items for each *conditional FP-Tree* . If it is not a single path, then *FP-Growth division is carried out* recursively.

The FP-Growth algorithm produces many frequent itemsets as the basic concept of association rules mining. The frequent itemsets are reviewed by quicksorting each itemset based on its support value to obtain a sequence of combinations that can be used as recommendations for business decisions [17].

Based on the description above, the author wants to utilize information technology by trying to design a program to help solve problems at the Asra Store by conducting further research with the title Application of the Fp-Growth Algorithm in Analyzing Consumer Purchasing Patterns in Asra Store Sales Transactions.

2. Research methodology

The research framework is the concept or stages that will be carried out in the research. So that the steps taken by the author in this design do not deviate from the main discussion and are easier to understand, the sequence of research steps will be made systematically so that it can be used as a clear and easy guide for solving existing problems. The research framework that the author carried out in the research can be depicted in Figure 1 .

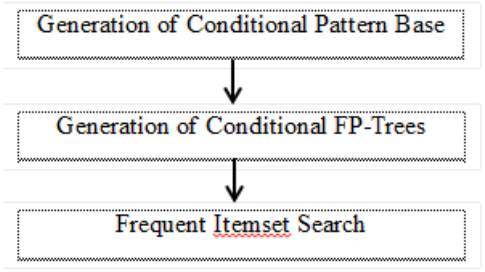

Figure 1. Research Framework

In this stage, a search will be carried out for a combination of items that meet the minimum requirements for *the support value* in the database. To get the *support value* for item A, you can use the following formula:

$$Support(A) = \frac{jumlah\ transaksi\ mengandung\ A}{total\ transaksi} \times 100\%$$

then , to get the *support value* of the two items, it is obtained using the following formula:

$$Support(A,B) = P(A \cap B) = \frac{jumlah\ transaksi\ mengandung\ A\ dan\ B}{total\ transaksi} \times 100\%$$

After all *frequent items* and *large item sets are obtained, the minimum confidence requirement (minconf)* can be found using the following formula:

$$Cofidence(A \rightarrow B) = P(A|B) = \frac{jumlah\ transaksi\ A\ dan\ B}{jumlah\ transaksi\ A} \times 100\%$$

3. Results and Discussion

3.1 System analysis

System analysis is a critical and very important stage in the development of information systems and aims to identify weaknesses that occur in current systems and old systems. Because analyzing the current system will make it easier to design and build the system to be built or a new system. By analyzing the old system, it can be used as a comparison, correction and development and design towards a new system.

In analyzing the system to be analyzed, the analyst must take several steps to understand all the procedures in the system. The steps in the system analysis stage are almost the same as the steps taken in defining system projects that will be developed in the system planning stage. The difference lies in the scope of the task.

3.2 Data analysis

Data analysis is everything that is obtained through observation or from a view. At this stage, data processing is carried out in accordance with the provisions of the formula that has been determined. Where in the calculation process the only data used in this research are sales transactions for November and December by limiting the 15 types of goods that are most frequently purchased and 10 transaction data as research objects. The following is *sample* data from transaction patterns with data calculations of 10 transaction data.

Table 1. Drug Sales Data

Transaction date	Product name
12/07/2021	Mineral Water, Coffee, Sugar, Tea, Candy
12/07/2021	Rice, Candy, Tea, Sugar, Coffee, Salt
12/07/2021	Bath Soap, Detergent, Shampoo, Mineral Water
12/07/2021	Eggs, Rice, Cooking Oil, Butter, Mineral Water
12/07/2021	Butter, Sugar, Mineral Water, Tea, Rice
12/07/2021	Toothpaste, Tea, Sugar, Eggs, Rice
12/07/2021	Cooking Oil, Rice, Sugar, Tea, Butter
12/07/2021	Tea, Mineral Water Sugar, Cooking Oil
12/07/2021	Candy, Tea, Sugar, Coffee, Shampoo, Mineral Water
12/07/2021	Toothpaste, Mineral Water, Bath Soap, Butter, Coffee, Milk, Sugar, Tea

3.3 . *Frequent Pattern Growth* Algorithm Calculation

Based on this data, the minimum requirement is 2 types of goods in 1 transaction with a minimum confidence value of 75% and minimum support of 20%. Before carrying out the stages of association rules using the Frequent Pattern Groth algorithm, in the first stage the data needed to compile the dataset is compiled. The results of the data collection are then transformed into a more specific form.

Table 2 . Results of Data Transformation Data Analysis

No	Product name	Product Code
1	Mineral water	KP001
2	Coffee	KP002
3	Sugar	KP003
4	Tea	KP004
5	Candy	KP005
6	Rice	KP006
7	Salt	KP007
8	Bath soap	KP008
9	Detergent	KP009
10	Shampoo	KP010
11	Egg	KP011
12	Cooking oil	KP012
13	Butter	KP013
14	Toothpaste	KP014
15	Milk	KP015

After carrying out the data transformation as in Table 4.1, product grouping is carried out from the transaction data, as follows:

Table 3. Grouped Transaction Data

Transaction Code	Sales Items/products
KT001	KP001, KP002, KP003, KP004, KP005
KT002	KP006, KP005, KP004, KP003, KP002, KP007
KT003	KP008, KP009, KP010, KP001
KT004	KP011, KP006, KP012, KP013, KP001
KT005	KP014, KP003, KP001, KP004, KP006
KT006	KP014, KP004, KP003, KP011, KP006
KT007	KP012, KP006, KP003, KP004, KP0013
KT008	KP004, KP003, KP001, KP012
KT009	KP005, KP004, KP003, KP010, KP001
KT010	KP014, KP001, KP008, KP013, KP002, KP015, KP003, KP004

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

Table 4. Frequent Itemset

No	Product	Frequency
1	KP003	8
2	KP004	8
3	KP001	7
4	KP006	5
5	KP013	4
6	KP002	3
7	KP005	3
8	KP012	3
9	KP008	2
10	KP010	2
11	KP011	2
12	KP014	2
13	KP007	1
14	KP009	1
15	KP015	1

After the highest and lowest frequencies are sorted, scan the data or rearrange the data that meets *minimum support* . For more details, see table 4.6, below.

Table 5. Dataset Sorted by Priority

No	Product
1	KP003,KP004,KP001,KP002,KP005
2	KP003,KP004,KP006,KP002,KP005
3	KP001, KP008, KP010
4	KP001,KP006,KP013,KP012,KP011
5	KP003,KP004,KP001,KP006,KP013
6	KP003,KP004,KP006,KP011,KP014
7	KP003,KP004,KP006,KP013,KP012
8	KP003, KP004, KP001, KP012
9	KP003,KP004,KP001,KP005,KP010
10	KP003,KP004,KP001,KP013,KP002,KP008,KP014

The results of creating a tree from transactions can be seen in Figure 2. In Figure 2, a frequent pattern tree is generated for all the products in Table 5

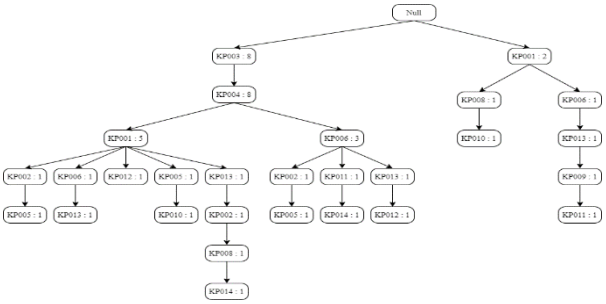

Figure 2 . FP-Growth Tree

Next, generate a *conditional pattern base* , as follows:

Table 6. Generation of Conditional Pattern Base

Product	Conditional Pattern Base
KP014	{KP003KP004KP001KP013KP002KP008:1}{KP003 KP004KP006KP011KP014:1}
KP011	{KP003KP004KP006:1}{KP001KP006KP013KP012: 1}
KP010	{KP003KP004KP001KP005:1}{KP001KP008KP010: 1}
KP008	{KP003KP004KP001KP013KP002:1}{KP001:1}
KP012	{KP003KP004KP001:1}{KP003KP004KP006KP013: 1}
KP005	{KP003KP004KP001KP002:1}{KP003KP004KP001: 1}{KP003KP004KP006KP002:1}
KP002	{KP003KP004KP001:1}{KP003KP004KP001KP013: 1}{KP003KP004KP006:1}
KP013	{KP003KP004KP001KP006:1}{KP003KP004KP001: 1}{KP003KP004KP006:1}{KP001KP006:1}
KP006	{KP003KP004KP001:1}{KP003KP004:3}{KP001:1}
KP001	{KP003KP004:5}
KP004	{KP003:8}
KP003	-

Next, generate a *conditional pattern base* , as follows:

Table 7. Generation of Conditional FP-Tree

Product	Conditional Pattern Base
KP014	<KP003:2,KP004:2>
KP011	<KP006:2>
KP010	<KP001:2>
KP008	-
KP012	<KP003:2,KP004:2>
KP005	<KP003:3,KP004:3,KP001:2,KP002:2>
KP002	<KP003:3,KP004:3,KP001:2>
KP013	<KP003:3,KP004:3,KP001:2,KP006:2>
KP006	<KP003:4,KP004:4>
KP001	<KP003:5,KP004:5>
KP004	<KP003:8>
KP003	-

Next, generate a *conditional frequent pattern tree* , as follows:

Table 8. Frequent Pattern Generation

Product	Frequent Pattern Generated
KP014	{KP003,KP014:2}{KP004,KP014:2}{KP003,KP00 4,KP014:2}
KP011	{KP006,KP011:2}
KP010	{KP001,KP010:2}
KP008	-
KP012	{KP003,KP012:2}{KP004,KP012:2}{KP003,KP00 4,KP012:2}
KP005	{KP003,KP005:3}{KP004,KP005:3}{KP002,KP00 5:2}{KP003,KP004,KP001,KP002,KP005:2}
KP002	{KP003,KP002:3}{KP004,KP002:3}{KP001,KP00 2:2}{KP003,KP004,KP001,KP002:2}
KP013	{KP003,KP013:3}{KP004,KP013:3}{KP001,KP01 3:3}{KP006,KP013:2}{KP003,KP004,KP001,KP00 6,KP013:2}
KP006	{KP003,KP006:4}{KP004,KP006:4}{KP003,KP00 4,KP006:4}
KP001	{KP003,KP001:5}{KP004,KP001:5}{KP003,KP00 4,KP001:5}
KP004	{KP003,KP004:8}
KP003	-

From the confidence calculation, there are 33 consumer habit patterns in shopping for products that meet the *confidence requirement* ≥ 0.75 , namely:

1. if consumers buy sugar, tea, mineral water, coffee, they will buy candy
2. If consumers buy sugar, tea, mineral water, candy, they will buy coffee with a confidence value of 1
3. If consumers buy sugar, coffee, candy, they will buy mineral water with a confidence value of 1
4. If consumers buy sugar, mineral water, coffee, candy, they will buy tea with a confidence value of 1
5. If consumers buy tea, mineral water, coffee, candy, they will buy sugar with a confidence value of 1
6. If consumers buy sugar, tea, mineral water, rice, they will buy butter with a confidence value of 1
7. If consumers buy sugar, tea, mineral water, butter, they will buy rice with a confidence value of 1
8. If consumers buy sugar, mineral water, rice, butter, they will buy mineral water with a confidence value of 1
9. If consumers buy sugar, mineral water, rice, margarine, they will buy tea with a confidence value of 1
10. If consumers buy tea, mineral water, rice, margarine, they will buy sugar with a confidence value of 1
11. If consumers buy sugar and toothpaste, they will buy tea with a confidence value of 1
12. If a consumer buys tea and toothpaste, he will buy sugar with a confidence value of 1
13. If a consumer buys toothpaste, he will buy tea and sugar with a confidence value of 1
14. If consumers buy sugar and cooking oil, they will buy tea with a confidence value of 1

15. If consumers buy tea and fried oil, they will buy sugar with a confidence value of 1
16. If consumers buy sugar and rice, they will buy tea with a confidence value of 1
17. If consumers buy tea and rice, they will buy sugar with a confidence value of 1
18. If consumers buy sugar and mineral water, they will buy tea with a confidence value of 1
19. If consumers buy tea and mineral water, they will buy sugar with a confidence value of 1
20. If a consumer buys toothpaste, he will buy sugar with a confidence value of 1
21. If a consumer buys toothpaste, he will buy tea with a confidence value of 1
22. If consumers buy eggs, they will buy rice with a confidence value of 1
23. if consumers buy shampoo, they will buy mineral water
24. If a consumer buys candy, he will buy sugar with a confidence value of 1
25. If a consumer buys candy, he will buy tea with a confidence value of 1
26. If consumers buy coffee, they will buy sugar with a confidence value of 1
27. If a consumer buys coffee, he will buy tea with a confidence value of 1
28. If consumers buy margarine, they will buy sugar with a confidence value of 0.75
29. if consumers buy margarine they will buy tea with a confidence value of 0.75
30. If consumers buy margarine, they will buy mineral water with a confidence value of 0.75
31. if consumers buy margarine, they will buy rice with a confidence value of 0.75
32. If a consumer buys sugar, he will buy tea with a confidence value of 1
33. If consumers buy tea, they will buy sugar with a confidence value of 1

4. Conclusion

The conclusions obtained from research in building a system for implementing the FP-Growth algorithm in analyzing consumer purchasing patterns at Toko Asra. After the various explanations and discussions that have been described in this research. From this, a conclusion can be drawn, namely that designing a data mining program using the FP-Growth method can help analyze consumer purchasing patterns based on goods purchased simultaneously. A web-based decision support system can process sales transaction data at Toko Asra into important information. Transaction data that is processed into information provides knowledge in calculating stock and strategic placement of goods in stores. The use of a decision support system is able to manage the stock system so that stock gaps no longer occur. The results of frequent itemset calculations can help find a sequence of combinations that can be used as product recommendations in business decisions.

References

[1] Fitriani, H. (2018). The contribution of fintech in increasing financial inclusion in agriculture (Analytical study using a sharia financial approach with peer to peer lending sites in agriculture in Indonesia). EL BARKA: Journal of Islamic Economics and Business, 1(1), 1-26. https://doi.org/10.21154/elbarka.v1i1.1392

[2] Yuliaty, T., Shafira, CS, & Akbar, MR (2020). MSME Strategy in Facing Global Business Competition. Mbia, 19(3), 293-308. https://doi.org/10.33557/mbia.v19i3.1167

[3] Siagian, AO, & Cahyono, Y. (2021). MSME marketing recovery strategy during the COVID-19 pandemic in the creative economy sector. Journal of Technology and Business Information Systems, 3(1), 206-217. https://doi.org/10.47233/jiteksis.v3i1.212

[4] Sianturi, RD (2020). Marketing management uses SWOT analysis for MSMEs to increase the competitiveness of MSMEs. Journal of Business and Economics Research (JBE), 1(1), 45-50. https://doi.org/10.47065/jbe.v1i1.118

[5] Wibowo, F., Khasanah, AU, & Putra, FIFS (2022). Analysis of the Impact of the Presence of Modern Markets on the Marketing Performance of Traditional Markets Based on the Perspectives of Traders and Consumers in Wonogiri Regency. Benefit: Journal of Management and Business (This Journal Has Migrated), 7(1), 53-65. https://doi.org/10.23917/benefit.v7i1.1605

[6] Fatihah, DC, & Saidah, IS (2021). Marketplace Promotion Model Based on Artificial Intelligence (AI) in Indonesia. JMBI UNSRAT (Scientific Journal of Business Management and Innovation, Sam Ratulangi University)., 8(3). https://doi.org/10.35794/jmbi.v8i3.35908

[7] Purnia, DS, & Warnilah, AI (2017). Implementation of Data Mining in Glasses Sales Using the Apriori Algorithm. IJCIT (Indonesian Journal on Computer and Information Technology), 2(2). https://doi.org/10.31294/ijcit.v2i2.2776

[8] Purba, KS, Hartama, D., & Suhada, S. (2022). Data Mining Analysis of the Distribution of Middle School Students in Pematangsiantar Using the K-Means Clustering Algorithm Method. Knight: Journal of Application of Information Systems (Computers and Management), 3(1), 1-8. https://doi.org/10.30645/kesatria.v3i1.91

[9] Artsitella, CR, Apriliani, AR, & Ashari, S. (2021). Application of association rules-market basket analysis to find frequent itemsets with the fp-growth algorithm. J. Al-AZHAR Indonesia. SCIENCE AND TECHNOLOGY SERIES, 6(2), 61.

[10] Purwaningsih, R., Widharto, Y., Susanto, N., & Utami, LT (2020). Redesign product layout in supermarkets based on purchasing behavior using the market basket analysis method. J@ti Undip: Journal of Industrial Engineering, 15(3), 196-202. https://doi.org/10.14710/jati.15.3.196-202

[11] Astrina, I., Arifin, MZ, & Pujianto, U. (2019). Application of the FP-Growth algorithm in determining consumer purchasing patterns for Medali Mas woven fabric. Matrix: Journal of Technology and Information Management, 9(1), 32-40. http://dx.doi.org/10.31940/matrix.v9i1.1036

[12] Deolika, A., Ginting, VS, Maryana, T., Sudiyarno, R., & Kusrini, K. (2019). Implementation of Apriori and Forward Chaining Algorithms to Determine the Right Food for Diabetes Sufferers. Journal of Information Technology, 3(2), 207-211. https://doi.org/10.36294/jurti.v3i2.1080

[13] Hasugian, B.S. (2019). Application of the Association Rule Method to Analyze Chemical Usage Patterns in the Laboratory Using the FP-Growth Algorithm (Case Study in the Chemical Laboratory of PT. PLN (Persero) Belawan Medan Generation Sector). Algorithms: Journal of Computer Science and Informatics, 3(2), 56. http://dx.doi.org/10.30829/algoritma.v3i2.6437

[14] Ghozali, MI, & Sugiharto, WH (2017). Shopping Pattern Analysis Using the Fp Growth Algorithm, Self Organizing Map (Som) and K Medoids. Symmetric: Journal of Mechanical Engineering, Electrical and Computer Science, 8(1), 317-326. https://doi.org/10.24176/simet.v8i1.995

[15] Achmad, F., Nurdiawan, O., & Wijaya, YA (2023). Analysis of Consumer Purchase Transaction Patterns in Health Retail Stores Using the FP-Growth Algorithm. JATI (Information Engineering Student Journal), 7(1), 168-175. https://doi.org/10.36040/jati.v7i1.6210

[16] Srisulistiowati, DB, Khaerudin, M., & Rejeki, S. (2021). Office Stationery Sales Prediction Information System Using the Fp-Growth Method (Case Study of the Bina Mulia School Cooperative Store). JSI (Journal of Information Systems) Suryadarma University, 8(2), 243-256. https://doi.org/10.35968/jsi.v8i2.739

[17] Harahap, S. Z., & Nastuti, A. (2019). Data Mining Technique for Determining Economical Packages for Basic Food and Daily Needs Using the Fp-Growth Algorithm (Case Study at Ulfamart Lubuk Alung). Informatics, 7(3), 111-119. https://doi.org/10.36987/informatika.v7i3.1381