ANALYSIS OF KIM 2 MEDAN PRODUCT SALES SYSTEM PATTERNS USING THE FP-GROWTH ALGORITHM

Zulham[1]*, Muhammad Eka [2], Hasan Khosasih [3]
Faculty of Engineering and Computer Science, Software engineering, Dharmawangsa University, Indonesia

Keywords:
Data Mining;
FP-Growth;
Algorithm;
Customers.

***Correspondence Address:**
zulham@dharmawangsa.ac.id

Abstract: Increasing revenue is one measure of a company's success that must be monitored. A company certainly hopes that there will be revenue growth every year which is stated in the revenue target. Companies must provide the best service strategy to markets and customers, but companies are always constrained by determining this strategy.
The aim of this research is to design a system using the FP-Growth method to recommend which products sell more/sold simultaneously in retail form at CV. Kharisma Jaya. The problem faced is that it is difficult to determine recommendations for products that are often sold or purchased simultaneously in retail form and the large number of products that are produced but whose sales are not yet optimal. Therefore, the FP-Growth method is here to provide a solution.
With this system, it is hoped that it can solve problems by making the FP-Growth method an alternative solution for companies in minimizing losses and increasing sales.

INTRODUCTION

The building materials industry is currently faced with the task of selecting the best building materials, one of which is CV.Kharisma Jaya, a practicing greenhouse industry specialist, designer and engineer in structures, site control, irrigation and planting. They offer not only economical products but also professional advice for a complete project wholesale center for all kinds of tools and greenhouse planting.

The industry certainly hopes for sales growth every year which is directed at sales targets. At certain times the level of income increases, and in other situations the opposite will happen. The same thing applies to CV.Kharisma Jaya, revenue is the most important thing for the company.

In the sales process, the seller or supplier of goods and services transfers ownership of the goods to the buyer at a certain price. Suppliers often experience problems because the desired product is not available or out of stock. This can result in consumers switching to other suppliers. Therefore, suppliers of goods must anticipate which products are in demand and frequently desired by current and future consumers. Several literatures show that there are several techniques that can be used to determine sales patterns, one of which is data mining.

Because in data mining there are ways and techniques to fulfill broad information needs and this information can be used as material for decision making. One way to apply data mining is by applying one of the association analysis functions, namely the FP-Growth algorithm.

RESEARCH METHODS

2.1. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) is the application of scientific methods to data mining. In this context, data mining is one step of the KDD process. Knowledge Discovery and Data mining (KDD) relates to integration techniques and scientific discovery, interpretation and visualization of patterns of data. Knowledge Discovery in Databases (KDD) is often used interchangeably to describe the process of extracting hidden information in a large database. Actually, these two terms have different terms, but are related to each other. And one of the stages in the entire KDD process is data mining.

2.2. Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. Data mining is a series of processes to explore added value from a collection of data in the form of knowledge that has not been known manually.

2.3. FP-Growth

Frequent pattern growth is a level of the Apriori association algorithm which uses alternative item set frequencies based on numbers that frequently appear in each transaction (frequent item sets) in a data set. The concept of the FP-Growth Algorithm is the formation of a tree or FP-Tree by searching for frequent item sets rather than using Apriori candidate generation in the algorithm process. By using this concept, the growth of the FP-Growth Algorithm is faster than the Apriori Algorithm. The FP-Growth growth algorithm is a further development of the Apriori algorithm.

2.4. FP-Growth

Association Rule is a Data Mining technique for finding associative rules between combinations of items. Association rules will use training data, according to the meaning of data mining, digging data to produce knowledge. Knowledge to find out which shopping items are often purchased simultaneously at one time. Association rules in the form of "if...then..." or "if...then..." are knowledge resulting from relationships based on data that has been extracted.

Association rules have two stages, as follows.

1. Look for frequent itemset values.
2. Determine the conditions and results of the search.

Association rules are obtained from the results of calculations consisting of 2 measures, namely:

1. Support

Support aims to find out how high the level of itemset dominance is in the transaction data that has been searched. The support value will determine the confidence value in whether an itemset is worth running or not. The support value can be used to find out the level of dominance of an item.

In determining the support value of an item, it can be obtained using the formula as below :

$$Support = \frac{NUMBER\ OF\ TRANSACTIONS\ CONTAINS\ A}{Total\ Transactions}\ x\ 100$$

$$Support\ A\ \cap B = \frac{NUMBER\ OF\ TRANSACTIONS\ CONTAINS\ A\ and\ B}{Total\ Transactions}\ x\ 100$$

2. Confidence

The confidence value is a relationship that shows the conditional relationship between two items :

$$Confidence\ = \mathrm{P}\ (\mathrm{A} \mid \mathrm{B}\ ) = \frac{NUMBER\ OF\ TRANSACTIONS\ CONTAINS\ A\ And\ \ B}{NUMBER\ OF\ TRANSACTIONS\ CONTAINS\ A}\ x\ 100$$

2.5. RapidMiner Studio

RapidMiner is an application that provides a unified environment for machine learning, deep learning, text mining, and predictive analytics. RapidMiner has around 500 data mining operators including input, output, data processing and visualization operators. RapidMiner is a standalone data analysis software and data mining engine that can be integrated into your own products.

RESULTS AND DISCUSSION

System testing aims to provide evidence that the input, process and output produced by RapidMiner Studio are correct and on target. Test the system by entering data into the system and recording the resulting output. If the input, process and output match, then the system is correct.

3.1. System Algorithm

The substance of this system algorithm is 3 things, namely: (1) flowchart of the solution used, (2) description of the data being tested, and (3) completion of the solution method or algorithm adopted.

The following is the data mining completion system algorithm. Data mining product recommendations on retail data on sales of imported building equipment products on CV. Kharisma Jaya :

1. FP-Growth Algorithm Flowchart
2. Determine the data to be processed
3. Generate Frequent Item sets
4. Addition of Transaction ID (TID)
5. Formation of FP-Tree
6. Sub Tree Formation
7. Association Rules.

Table 1. Transaction Data

TID	Product name
1	Ventilation System,Heating system, Shading System
2	NoozleAgridor, Threaded Nut, Dinamic sprayer
3	Plastic Sheet, Plastic Fasteners, Dinamic sprayer
4	Shade Scrolling Tube, Heating system, Shading System, Dinamic sprayer, Film Rolling System
5	Threaded Nut, Dinamic sprayer, Shade Scrolling Tube, Heating system

6	Shade Scrolling Tube, Heating system, Shading System, Dinamic sprayer, Threaded Nut
7	Plastic Sheet, Plastic Fasteners, Dinamic sprayer, Heating system, Threaded Nut, NoozleAgridor
8	NoozleAgridor, Threaded Nut, Dinamic sprayer, Shade Scrolling Tube, Heating system
9	Ventilation System, Heating system, Shading System, Shade Scrolling Tube
10	NoozleAgridor, Threaded Nut, Dinamic sprayer, Plastic Sheet, Plastic Fasteners
11	Shade Scrolling Tube, Heating system, Shading System, Dinamic sprayer, Film Rolling System, Threaded Nut,
12	Threaded Nut, Dinamic sprayer, Shade Scrolling Tube, Heating system, Shading System
13	Heating system, Shading System, Film Rolling System
14	Threaded Nut, Shade Scrolling Tube, Heating system
15	Plastic Sheet, Plastic Fasteners, Dinamic sprayer
16	NoozleAgridor, Threaded Nut, Dinamic sprayer, Shade Scrolling Tube
17	Film Rolling System, NoozleAgridor, Threaded Nut, Heating system
18	Dinamic sprayer, Shade Scrolling Tube, Heating system
19	Heating system, Threaded Nut, NoozleAgridor, Shade Scrolling Tube
	TID 20 to 50
51	Plastic Sheet, Plastic Fasteners, Dinamic sprayer, Heating system
52	Heating system, Shading System, Plastic Fasteners, Dinamic sprayer, Shade Scrolling Tube
53	Plastic Fasteners, Dinamic sprayer, Heating system
54	Plastic Sheet, Plastic Fasteners, Dinamic sprayer, Heating system, Threaded Nut, NoozleAgridor
55	Dinamic sprayer, Threaded Nut, Ventilation System, Heating system
56	Heating system, Threaded Nut, Plastic Sheet, Shade Scrolling Tube
57	Heating system, Shading System, Dinamic sprayer, Threaded Nut
58	Film Rolling System, NoozleAgridor, Threaded Nut, Heating system
59	Plastic Sheet, Plastic Fasteners, Dinamic sprayer
60	Shade Scrolling Tube, Heating system, Shading System, Dinamic sprayer, Film Rolling System

To make the process easier, the transaction data table will be replaced with a code. The following is a table for changing product names to codes

Table 2. Product Code

Code	Product Name
A01	Ventilation System
A02	Heating system
A03	Shading System
A04	Noozle Agridor
A05	Threaded Nut
A06	Dinamic sprayer

A07	Plastic Sheet
A08	Plastic Fasteners
A09	Shade Scrolling Tube
A10	Film Rolling System

The frequency and support for each item are sorted from the highest and then the item support value is searched for from 60 transactions. Then you will get the support value as in the following table:

Table 3. Frequency of Appearance of Each Item

Item	Frequency	Support
A01	9	(9/60)x100% = 15%
A02	43	(43/60)x100% = 71,667%
A03	18	(18/60)x100% = 30%
A04	18	(18/60)x100% = 30%
A05	37	(37/60)x100% = 61,667%
A06	41	(41/60)x100% = 68,333%
A07	20	(20/60)x100% = 33,333%
A08	18	(18/60)x100% = 30%
A09	31	(31/60)x100% = 51,67%
A10	13	(13/60)x100% = 21,667%

Based on the table above which contains the support value for each item, the minimum support value is set = 40%. The following is a table that meets the minimum support value = 40%.

Table 4. Frequency of Appearance of Each Item

Item	Frequency	Support
A02	43	71,667%
A06	41	68,333%
A05	37	61,667%
A09	31	51,67%

From the value table containing the support value for each item, it will be sorted based on the highest frequency of appearance with a minimum support value = 40%. Below is a table of transaction sequence data based on highest support :

TID depiction is carried out up to the last TID. The following are the results of the FP-Tree depiction for TID 60.

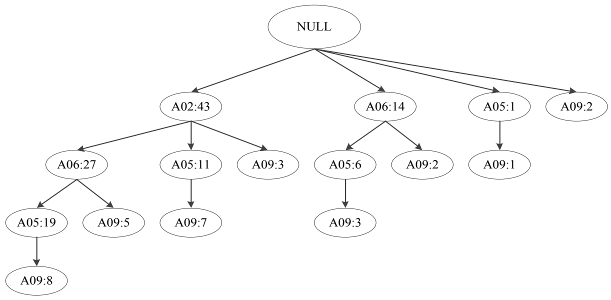

Figure 1. FP-Tree formation results after reading TID 60

Based on the image above, the priority order for the item with the smallest count is A09 with a count value of 31 and the highest count is A10 with a count value of 10. Once the smallest count is known, a subtree will be created ending in node A21. Conditional pattern base, conditional FP-Tree and frequent item sets will be found from the subtree.

1. Sub Tree A09

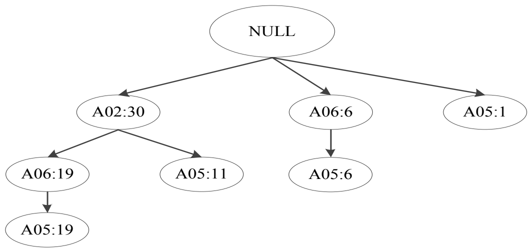

Figure 2. The path containing node A09

a. Conditional Pattern Base

A09: {A02, A06, A05 : 8}, {A02, A06 : 5}, {A02, A05 : 7 }, {A02 : 3}, {A06, A05 : 3}, {A06 : 2}, {A05 : 1}

The conditional pattern base is obtained by reading each path ending in node A09.

b. Conditional FP-Tree :

After the conditional pattern base is obtained, the conditional FP-Tree is formed by ignoring the single item on the path ending in node A09.

A09 : {A02 : 23}, {A06 : 18}, {A05 : 19}

c. Frequent Itemsets:

Next, to get frequent item sets, combine the items that will create a conditional FP-Tree provided that the count of the items meets the minimum support.

A09: {A02, A09 : 23}, {A06, A09 : 18}, {A05, A09 : 19}

2. Sub Tree A05

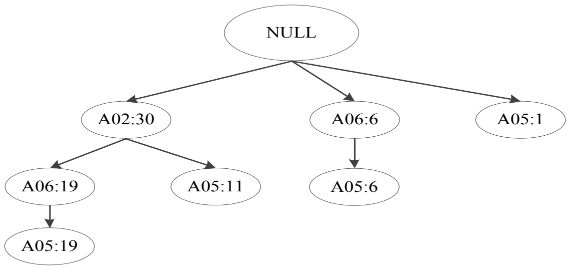

Figure 3. The path containing node A05

a. Conditional Pattern Base:

A05: {A02, A06: 19}, {A02 : 11}, {A06 : 6}

The conditional pattern base is obtained by reading each path ending in node A05.

b. Conditional FP-Tree :

After the conditional pattern base is obtained, the conditional FP-Tree is formed by ignoring the single item on the path ending in node A05.

A05 : {A02 : 30}, {A06 : 6}

c. Frequent Itemsets:

Next, to get frequent item sets, combine the items that will be made into a conditional FP-Tree provided that the count of the items meets the minimum support.

A09: {A02, A05 : 30}, {A06, A05 : 25}

3. Sub Tree A06

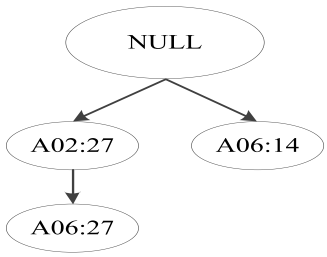

Figure 4. The path containing node A06

a. Conditional Pattern Base

A06: {A02 : 27}

The conditional pattern base is obtained by reading each path ending in node A06.

b. Conditional FP-Tree :

After the conditional pattern base is obtained, the conditional FP-Tree is formed by ignoring the single item on the path ending in node A06.

A06 : {A02 : 27}

c. Frequent Itemsets:

Next, to get frequent item sets, combine the items that will be created as a conditional FP-Tree provided that the count of the items meets the minimum support.

A06: {A02, A06 : 27}

Sub tree drawing is done for all item sets. Based on the results of the sub tree that has been formed, a subset table will be formed with 2 item combinations.

The next process will be calculated because it has fulfilled the frequent item set requirements to produce an Association Rule which has at least 2 items where if you open category A then you will open category B. The following is a subset that is eligible to calculate its confidence level.

Table 5. Subset

TID	Frequent Itemsets	Subsets
1	A09	{A02, A09 : 23}, {A06, A09 : 18}, {A05, A09 : 19}
2	A05	{A02, A05 : 30}, {A06, A05 : 25}
3	A06	{A02, A06 : 27}

After finding a subset that meets the requirements, the next frequency value will be obtained according to the subset. The following is a table of frequent pattern.

Table 6. Frequent Pattern

TID	Frequent Itemsets	Subsets
1	{A02, A09}	23
2	{A06, A09}	18
3	{A05, A09}	19
4	{A02, A05}	30
5	{A06, A05}	25
6	{A02, A06}	27

At this stage, calculations will be carried out to determine the support value for each item set.

Table 7. Frequent Support Association Rules

TID	Frequent Itemsets	Subsets	Support
1	{A02, A09}	23	(23/60) x 100% = 38,333%
2	{A06, A09}	18	(18/60) x 100% = 30%

3	{A05, A09}	19	(19/60) x 100% = 31,667%
4	{A02, A05}	30	(30/60) x 100% = 50%
5	{A06, A05}	25	(25/60) x 100% = 41,667%
6	{A02, A06}	27	(27/60) x 100% = 45%

After knowing the results of the Support value calculation, it will then be eliminated according to the minimum support = 40%. The following are the results of the elimination of 2 item sets.

Table 8. Eliminate Support Association Rules

TID	Frequent Itemsets	Subsets	Support
1	{A02, A05}	30	(30/60)x100% = 50%
2	{A06, A05}	25	(25/60)x100% = 41,667%
3	{A02, A06}	27	(27/60)x100% = 45%

After going through the process of eliminating the support values for the 2 item sets, calculations are then carried out to determine the confidence value for each item set. Next, you will get a confidence value like the following table:

Table 9. Frequent Confidence Association rules

TID	Frequent Itemsets	Subsets	Confidence
1	{A02, A05}	30	(30/43) x 100% = 69,767%
2	{A06, A05}	25	(25/41) x 100% = 60,976%
3	{A02, A06}	27	(27/43) x 100% = 62,791%

From the various stages that have been carried out previously, the association rules are :

1. If you buy a Heating system (A02) you will buy a Threaded Nut (A05) with a support value of 50% and confidence of 69.767%.
2. If you buy a Dynamic sprayer (A06), you will buy a Threaded Nut (A05) with a support value of 41.667% and confidence of 60.976%.
3. If you buy a Heating system (A02), you will buy a Dynamic sprayer (A06) with a support value of 45% and confidence of 62.791%.

CONCLUSION

Based on research that has gone through the design and evaluation stages of implementing the FP-Growth Algorithm for Product Recommendations on Retail Data on Sales of Imported Building Equipment Products at (CV.Kharisma Jaya) it can be concluded as follows:

1. CV.Kharisma Jaya is a private company that has a policy of recommending products that are more frequently sold or purchased simultaneously in retail form.
2. Based on the research conducted, the FP-Growth method can be applied as a strategy to increase sales of building materials.
3. Based on the research results, the system that has been designed can be used as a solution to recommend products sold at retail precisely and accurately.
4. The system that has been designed is then tested and implemented with RapidMiner entering data according to previous transactions, the output results are in accordance with manual calculation data.

REFERENCE

D. Hartanti and V. Atina, "Product Stock Supply Analysis System with FP Growth Algorithm," vol. 5, no. 4, pp. 1312–1320, 2023, doi: 10.51519/journalisi.v5i4.580.

W. Hadinata, J. Waruwu, and T. Hermanto, "Journal of Sisfotek Global Comparison of Apriori and Frequent Pattern Growth Algorithm in Predicting The Sales of Goods ARTICLE HISTORY," Issn, vol. 11, no. 2, pp. 89–96, 2021, [Online]. Available: http://journal.stmikglobal.ac.id/index.php/sisfotek

M. Hafizh, T. Novita, D. Guswandi, H. Syahputra, and L. Mayola, "Implementasi Data Mining Menggunakan Algoritma Fp-Growth Untuk Menganalisa Transaksi Penjualan Ekspor Online," J. Teknol. Dan Sist. Inf. Bisnis, vol. 5, no. 3, pp. 242–249, 2023, doi: 10.47233/jteksis.v5i3.847.

R. L. Najmi, M. Irsyad, F. Insani, A. Nazir, and . P., "Analisis Pola Asosiasi Data Transaksi Penjualan Minuman Menggunakan Algoritma FP-Growth dan Eclat," Build. Informatics, Technol. Sci., vol. 5, no. 1, pp. 0–7, 2023, doi: 10.47065/bits.v5i1.3592.

D. Agushinta R. and M. M. Putri, "Association Rule Analysis of Fp-Growth Algorithm on

Drug Purchase Patterns," J. Ilm. Teknol. dan Rekayasa, vol. 27, no. 3, pp. 196–212, 2022, doi: 10.35760/tr.2022.v27i3.4626.

AHMAD ADRI, "Implementasi Data Mining Menggunakan Algoritma Apriori," Pap. Knowl. . Towar. a Media Hist. Doc., vol. 6, no. 2, pp. 1–77, 2021.

A. Wadanur and A. A. Sari, "Implementasi Algoritma Apriori dan FP-Growth pada Penjualan Spareparts," Edumatic J. Pendidik. Inform., vol. 6, no. 1, pp. 107–115, 2022, doi: 10.29408/edumatic.v6i1.5470.

A. H. Nasyuha et al., "Frequent pattern growth algorithm for maximizing display items," Telkomnika (Telecommunication Comput. Electron. Control., vol. 19, no. 2, pp. 390–396, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16192.

L. N. Rani, S. Defit, and L. J. Muhammad, "Determination of Student Subjects in Higher Education Using Hybrid Data Mining Method with the K-Means Algorithm and FP Growth," Int. J. Artif. Intell. Res., vol. 5, no. 1, pp. 91–101, 2021, doi: 10.29099/ijair.v5i1.223.

D. K. Pramudito, T. W. Nurdiani, B. Winardi, A. Y. Rukmana, and K. Kraugusteeliana, "Website User Interface Design Using Data Mining Task Centered System Design Method At National Private Humanitarian Institutions," Indones. J. Artif. Intell. Data Min., vol. 6, no. 2, p. 281, 2023, doi: 10.24014/ijaidm.v6i2.25814.

I. Riadi, H. Herman, F. Fitriah, and S. Suprihatin, "Optimizing Inventory with Frequent Pattern Growth Algorithm for Small and Medium Enterprises," MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 23, no. 1, pp. 169–182, 2023, doi: 10.30812/matrik.v23i1.3363.

N. Suhandi and R. Gustriansyah, "Marketing Strategy Using Frequent Pattern Growth," J. Comput. Networks, Archit. High Perform. Comput., vol. 3, no. 2, pp. 194–201, 2021, doi: 10.47709/cnahpc.v3i2.1039.

E. E. Putri, B. S. Hasugian, A. Info, and D. Mining, "Pattern Analysis of Drug Procurement System With FP-Growth Algorithm," vol. 7, no. 1, pp. 70–79, 2022, doi: 10.15575/join.v7i1.841.

Y. P. Bunda, "Algoritma FP-Growth Untuk Menganalisa Pola Pembelian Oleh-Oleh (Studi Kasus Di Pusat Oleh-Oleh Ummi Aufa Hakim)," Riau J. Comput. Sci., vol. 06, no. 01, pp. 34–44, 2020.

E. Munanda and S. Monalisa, "Penerapan Algoritma Fp-Growth Pada Data Transaksi Penjualan Untuk Penentuan Tataletak," J. Ilm. Rekayasa dan Manaj. Sist. Inf., vol. 7, no. 2, pp. 173–184, 2021, [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/13253

G. Guntoro and C. P. Hutabarat, "Penerapan Data Mining Association Rule Menggunakan Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel," J. Komtika (Komputasi dan Inform., vol. 5, no. 2, pp. 112–121, 2021, doi: 10.31603/komtika.v5i2.6251.

A. H. Nasyuha, Zulham, and I. Rusydi, "Implementation of K-means algorithm in data analysis," Telkomnika (Telecommunication Comput. Electron. Control., vol. 20, no. 2, pp. 307–313, 2022, doi: 10.12928/TELKOMNIKA.v20i2.21986.

F. Mahardika, N. Alfiah, and R. Bagus Bambang Sumantri, "Penerapan Metode FP Tree dan Frequent Pattern Growth pada Penerimaan Mahasiswa Baru STMIK," Blend Sains J. Tek., vol. 1, no. 3, pp. 226–234, 2023, doi: 10.56211/blendsains.v1i3.176.

Q. D. Rosyadi and A. W. Utami, "Implementasi Algoritma Frequent Pattern Growth untuk Menentukan Pola Pembelian Konsumen pada Toko Tanaman Berbasis Website," vol. 04, no. 03, pp. 107–114, 2023.

I. M. D. P. Asana, I. K. A. G. Wiguna, K. J. Atmaja, and I. P. A. Sanjaya, "FP-Growth Implementation in Frequent Itemset Mining for Consumer Shopping Pattern Analysis Application," J. Mantik, vol. 4, no. 3, pp. 2063–2070, 2020, [Online]. Available: http://iocscience.org/ejournal/index.php/mantik/article/view/882/595.