

Analysis of Sales by Using Apriori and FP-Growth at PT. Panca Putra Solusindo

¹Sita Anggraeni, ²Marlina Ana Iha, ³Wati Erawati, ⁴Sayyid Khairunnas
STMIK Nusa Mandiri, STMIK Nusa Mandiri, Universitas Bina Sarana Informatika,
Universitas Bina Sarana Informatika
Jakarta, Indonesia

sita.sia@nusamandiri.ac.id, marlina.anaiha@yahoo.co.id, wati.wti@bsi.ac.id, sayyid.skh@bsi.ac.id

Abstract— PT. Panca Putra Solusindo is a company which sells Electronic equipment. It has a lot of transactions. Among these data, the company has not been able to provide such platform that can be used as information for management to find out which products are most preferred by consumers. Data mining is present to provide information patterns to companies in answering these needs and helping in marketing to be more effective. In this research, the function of the Apriori algorithm association is used to find out the minimum support and minimum confidence and FP-Growth in knowing the frequent itemset and both of them can find out the best rule that occurs in the transaction. These two algorithms are tested by using Weka application version 3.8.

Keywords: *Data Mining, Association Rule, Apriori, Frequent pattern growth, FP-Growth, WEKA*

I. INTRODUCTION (HEADING 1)

The increase of business competition requires developers to find a strategy that can enhance sales and marketing of products, one of which is to use sales data. With sales activities every day, the data will grow more and more. The data does not only function as an archive for the company, the data can be utilized and processed into useful information for increasing sales of the product and promotions.

Electronic products are items that are very much needed today, because electronic equipments are very helpful for humans in carrying out various activities. This study raised the problems that exist in PT. Panca Putra Solusindo, which is still not providing the most sold and interconnected products. Then the promotion of products that are made in one interconnected package is not maximized. In addition, there are still many products in the inventory, but the sales process is not or has not been maximal.

One method that can be used is by applying data mining usage. Because in data mining, there

are ways and techniques in meeting broad information needs and that information can be used as material for decision making. The implementation of data mining is by performing one of the association functions in data mining using apriori algorithms and analyst associations, namely FP-Growth algorithm. The Apriori algorithm which aims to find frequent itemsets is run on a set of data. Apriori analysis defined a process for finding all apriori rules that meet the minimum requirements for support and minimum requirements for confidence. As for FPGrowth itself, it is an alternative algorithm that can be used in terms of determining the set of data that most frequently appears (frequent itemset) in a set of data.

According to those backgrounds, the statements of problems as follows:

1. How to apply the apriori and FP-Growth algorithms to find out the sales of the most sold electronic products?
2. Can apriori algorithms and FP-Growth help to develop marketing strategies?

3. How to implement apriori and FP-Growth algorithms on sales with the Weka application?

The purpose of this research is: To determine the extent to which apriori and FP-Growth algorithms can help the development of marketing strategies with the implementation of Weka's data mining application software.

The benefit of this research is: Helping companies to find out the most sold electronic products.

II. LITERATURE REVIEW

Data mining is a term used to describe certain knowledge and discover data in a database. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and learning machine to extract and identify useful information and related knowledge from various large databases (KusriniEmha, 2010)

Data mining has several main functions which in each function has different usage goals, namely

A. Description, to give a brief description of a large number of data sets and types.

B. Estimation, to guess an unknown value, for example guessing someone's income when information about that person is known.

C. Prediction, to estimate future value, for example predicting the stock of goods in the next year.

D. Classification, is a process of finding a model or function that explains or distinguishes a concept or class of data, with the aim of being able to estimate the class of an object whose label is unknown.

E. Clustering, grouping and identifying data that has certain characteristics.

F. Association, also called market basket analysis where this function identifies products that consumers might purchase together with other products.

Apriori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 to determine Frequent itemsets for Boolean association rules. Apriori algorithms include the types of Association Rules for data mining. Rules that state associations between several attributes are often called affinity analysis or market basket analysis. Association analysis or association rule mining is a data mining technique to find the rules of an item combination. One of the stages of association analysis that attracted the attention of many researchers to produce efficient algorithms is the analysis of high frequency patterns. The importance of an association can be known by

two benchmarks, namely: support and confidence. Support (supporting value) is the percentage of the combination of items in the database, whereas confidence (the value of certainty) is the strength of the relationship between items in the rules of association (KusriniEmha, 2010)

1. Analysis of High Frequency Patterns with Apriori Algorithms

This stage looks for item combinations that meet the minimum requirements of the support value in the database. The support value of an item is obtained by using the following formula:

$$\text{support}(A) = \frac{\text{Jumlah transaksi mengandung } A}{\text{Total transaksi}}$$

Meanwhile, the support value of 2 items is obtained using the formula:

$$\text{support}(A, B) = P(A \cap B)$$

$$\text{support} = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\sum \text{transaksi}}$$

The frequent itemset shows itemset which have a frequency of occurrence of more than the specified minimum value (ϕ). For instance $\phi = 2$, then all itemsets whose frequency of occurrence is more than or equal to 2 times are called frequent. The set of frequent k-itemset is denoted by F_k .

2. Formation of Association Rules

After all the high frequency patterns are found, then association rules are sought that meet the minimum requirements for confidence by calculating the associative confidence rule A B. The confidence values of the A B rule are obtained by the following formula:

$$\text{confidence} = \frac{\sum \text{transaction containing } A \text{ and } B}{\sum \text{transaction containing } A}$$

To determine the association rules to be chosen, they must be sorted according to the Support \times Confidence. Rules are taken as many as n rules that have the greatest results.

The apriori algorithm is divided into several stages called narration or pass

1. Forming itemset candidates. The k-itemset candidate is formed from the combination (k-1) itemset obtained from the previous iteration. One way of apriori algorithm is to trim the k-itemset candidate

whose subset contains k-1 items not included in the high-frequency pattern with length k-1.

2. Calculation of support for each k-itemset candidate. Support from each k-itemset candidate is obtained by scanning the database to calculate the number of transactions that contain all items in the k-itemset candidate. This is also a feature of apriori algorithm where it is necessary to calculate the entire database as many as the longest k-itemset.

3. Set a high frequency pattern.

The high frequency pattern that contains k items or k-itemset is determined from the k-itemset candidate whose support is greater than the minimum support.

4. If no new high frequency pattern is obtained then the whole process is stopped.

The FP-Growth algorithm is a development of the Apriori algorithm. Frequent Pattern Growth (FP-Growth) algorithm is an alternative algorithm that can be used to determine the frequent itemset in a data set [4].

The FP-Growth algorithm uses the concept of tree development, commonly called FP-Tree, in frequent itemset search instead of generate candidate as done in the Apriori algorithm. By using this concept, the FP-Growth algorithm is faster than the Apriori algorithm.

The FP-Growth algorithm has stages that must be passed in order to provide maximum results, these stages are:

1. The stage of generating conditional pattern base.
2. The stage of generating conditional FP-Tree.
3. The Stage of frequent itemset search.

Association rules is a process in data mining to determine all associative rules that meet the minimum requirements for support (minsup) and confidence (minconf) in a database. Both of these conditions will be used for applying interesting association rules compared to the prescribed limits, namely minsup and minconf. In determining the minimum support value of an item, it can be obtained by using the formula (1) as below:

$$\text{Support}(A) = \frac{\text{Amount of Transactions Containing } A}{\text{Total Transactions}}$$

$$\text{Support}(A \cap B) = \frac{\text{Amount of Transactions Containing } A \text{ and } B}{\text{Total Transactions}}$$

Whereas in determining minimum confidence can be determined by the formula equation (2) as below:

$$\text{Confidence}(A \rightarrow B) = P(A|B) = \frac{\text{Amount of Transactions Containing } A \text{ and } B}{\text{Amount of Transactions Containing } A}$$

III. PROPOSED METHOD

In composing this study, transaction data, in which sales data, is taken from the sale of PT. Panca Putra Solusindo Jakarta in January 2016 to December 2016, namely the association of data mining using the Apriori Algorithm and FP-Growth methods. The authors took a sample of 116 data from the total population of data on sales transaction data of PT Panca Putra Solusindo Jakarta, which amounted to 130 data. The sample to be taken is up to 116 top data contained in the primary data and contains 14 items.

Before analyzing the Association of data mining (Moh SholikAbu, 2018), we need the Apriori algorithm flowchart as follows :

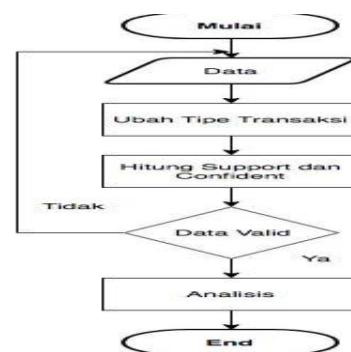


Figure of Flowchart of Apriori Algorithms

Then the steps in determining data mining associations using the FP-Growth algorithm are as follows :

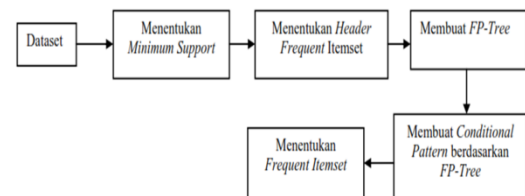


Figure of Block Image of FP-Growth Algorithm Diagram

Furthermore, the author will prepare data with tabulations in the form of excel with the terms "Y" and "N" to describe that buying and noting an item of

goods (SensuseGoldie, 2012), then the data is changed in the form of "CSV". The following tabulation data is as follows.

Table of Tabulation of Transaction data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	M Transist	Keyboard	DICRT	Leno	Memory	Monitor HP	NB Dell	NB HP	NB Lenovo	Electro	PC Dell	PC HP	PC Lenovo	Printer HP	Scanner HP	Flashdisk
2	Y	N	N	N	N	N	N	Y	Y	N	Y	N	N	N	N	N
3	2	N	N	N	N	N	N	Y	Y	N	N	N	N	N	N	N
4	3	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N
5	4	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N
6	5	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N
7	6	N	N	N	N	N	N	N	N	Y	N	N	Y	N	N	Y
8	7	N	N	N	N	Y	N	N	Y	N	N	N	N	N	N	N
9	8	N	N	N	N	N	N	N	Y	N	N	Y	Y	Y	N	N
10	9	N	N	N	Y	N	N	N	Y	N	N	Y	Y	Y	N	N
11	10	N	N	N	N	N	N	N	N	Y	Y	Y	N	N	N	Y
12	11	Y	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N
13	12	Y	N	N	N	N	N	Y	N	Y	N	Y	Y	N	N	N
14	13	N	N	N	N	N	N	Y	N	N	Y	N	Y	N	N	N
15	14	N	N	N	N	N	N	Y	N	N	N	Y	Y	N	N	N
16	15	Y	N	N	N	Y	N	N	N	N	N	Y	N	N	N	N
17	16	Y	N	N	N	N	N	N	Y	Y	Y	N	N	Y	N	Y

The next step is to transform the coding into the Weka data mining application (HagugianPaska, 2017) by taking part in the Association with the Apriori and FP-Growth methods.

IV. RESULT AND DISCUSSION

The important phase of data mining analysis Apriori algorithm is determining the highest frequency pattern by calculating the support value of one set item, so that the table is formed as follows:

Tabel of Support Per Item

Item	Support	Support (%)
PC HP	53	46%
Notebook HP	35	30.2%
PC Lenovo	32	27.6%
Notebook Lenovo	29	25%
Memory	27	23.3%
Other Electronic	21	18.10%
USB Flashdisk	17	15%
PC Dell	16	14%
Printer HP	16	14%
Monitor	7	6.03%
Keyboard	6	5.2%
LCD/CRT Lenovo	5	4.31%
Notebook Dell	3	2.6%
Scanner HP	1	0.9%

From the table above, minimum support can be determined at 9%, then itemset will be determined which can meet the minimum support as follows:

Table of Itemset that meets the minimum support

Item	Support	Support (%)
PC HP	53	46%
Notebook HP	35	30.2%
PC Lenovo	32	27.6%
Notebook Lenovo	29	25%
Memory	27	23.3%
Other Electronic	21	18.10%
USB Flashdisk	17	15%
PC Dell	16	14%
Printer HP	16	14%

Next phase, two item frequency pattern formation is formed from items of the type of goods that meet the minimum support by combining all items into two combinations, the result of the combination of two items results in a minimum support of 9%. Then the items that meet the requirements are as follows:

Table of two itemset that fulfill minimum support

Item	Support	Support (%)
PC HP x Notebook HP	13	11.20%
Notebook HP x Notebook Lenovo	12	10.34%
Notebook HP x Memory	11	9.5%
PC HP x PC Lenovo	10	9%
PC HP x Other Electronic	10	9%
PC HP x PC Dell	10	9%

After the minimum support value is formed, the author determines how to determine the value of confidence. In this case, it will be tested using the Weka application.

The writers use Weka application then select explore to see the "CSV" type data then by using the Apriori algorithm the Association rules will be formed as many as ten best rules with the determination of confidence of 0.4 so that the image looks as follows:

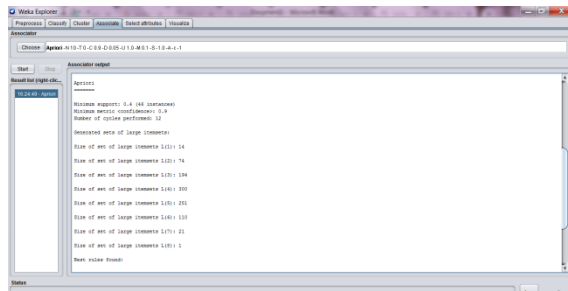


Figure of Minimum Confidence in Apriori using Weka Application

From the results above, we will get some rules in the apriori method as follows:

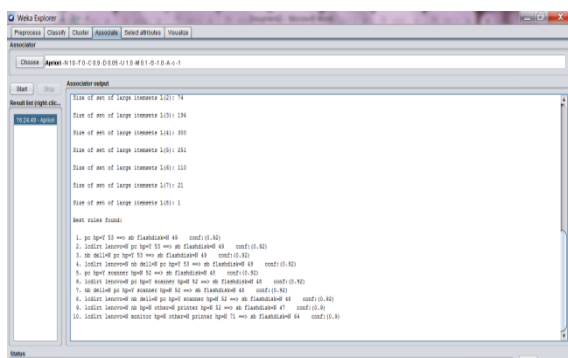


Figure of Best Rule Apriori with Weka Application

And items that have frequent frequent items are PC HP for 53 times the same sales as the following picture:

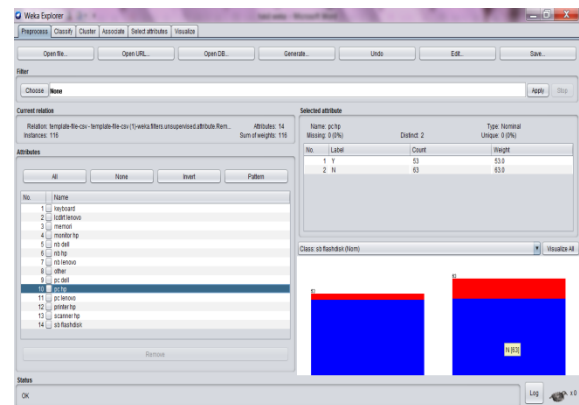


Figure of Frequent item Apriori with Weka Application

After being proven by testing the Weka application, the next phase is data processing

based on the FP-Growth Algorithm, in this case the data is changed back in one format, namely the condition "true" or "t" as shown below:

Table of FP-Growth Tabulation

ID	Keyboard	DCRT Lens	Memory	Monitor	HP Dell	HP HP	HP Lenovo	Electro	PC Dell	PC HP	PC Lenovo	Printer HP	Scanner HP	HP HP
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														
19														

After tabulation as shown above, the data is prepared in the format "CSV", then open the Weka application to calculate the value of confidence and the rule produced.

In processing data here, it is necessary to convert Weka application data into NominaltoBinary and NumeriktoBinary in the Unsupervised Attribute filter format as shown below:

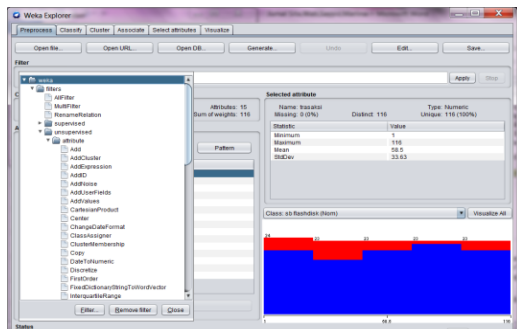


Figure of Filter Type FP-Growth data

Then after the data type filter is appropriate, rules will be obtained by determining the minimum confidence 0.4, which is as follows:

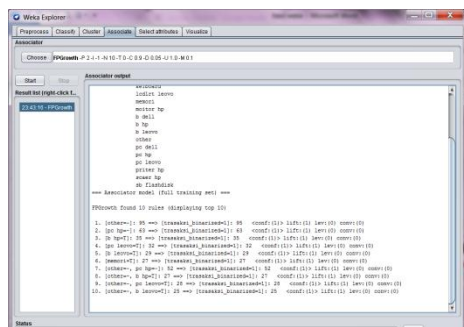


Figure of of FP-Growth results in Weka Application

Overall, both the Apriori Association algorithm and the expansion of FP-Growth produce several rules that often occur in a transaction (frequent items). This helps management to expand marketing strategies in increasing sales according to the combination of items in the association method that has been described and tested by the Weka application.

V. CONCLUSION AND SUGGESTION

1. In the Apriori Algorithm, the confidence value of an item is 92% and we can clearly know what products are sold high in a period, while the FP-Growth algorithm produces several rules that have the most value, namely frequent itemset, so they can be rooted in making the next FP-Tree.
2. The best rules are produced from the Apriori and FP-Growth algorithms which provide

several choices with a predetermined minimum confidence value of 0.4 so that it can provide data to marketing to find out in the future in the sales strategy planning to improve a better performance.

3. The author is aware of many shortcomings in data processing and data retrieval, the authors hope that in the future there will be other writers developing different algorithmic methods and producing accurate analysis for the needs of the company.

VI. ACKNOWLEDGMENT

We gratefully thank to the principal of PT Panca Putra Solusindo for allowing our to conduct the research there. We thank our colleagues from STMIK Nusa Mandiri and Universitas Bina Sarana Informatika who provided insight and expertise that greatly assisted the research.

VII. REFERENCES

- Hagugian, P. M. (2017). Pengujian Algoritma Apriori dengan Aplikasi Weka dalam Pembentukan Asosiasi Rule. *Jurnal Mantik Penusa*, 98-103.
- Kennedi Tampubolon, H. S. (2013). Implementasi Data Mining Algoritma Apriori pada Sistem Persediaan Alat-alat Kesehatan. *Majalah Ilmiah INTI*, 93-106.
- Khoiriah, R. Y. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Citec Journal*, 102-113.
- Kusrini, E. T. (2010). Algoritma Data Mining. In E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi.
- Moh Sholik, A. S. (2018). Implementasi Algoritma Apriori untuk mencari Asosiasi Barang yang dijual di E-Commerce Ordermas. *Techno Com*, 158-170.
- Nugroho Wandu, R. A. (2012). Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalan Association Rule menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur). *Jurnal Teknik ITS*, 445-449.

- Purnamasari, D. (2013). Get Easy Using Weka. In D. Purnamasari, *Get Easy Using Weka*. Jakarta: Dapur Buku.
- Samuel, D. (2008). Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset. In D. Samuel, *Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset*. Bandung: Institut Teknologi Bandung.
- Sensuse, G. G. (2012). Penerapan Metode Data Mining Market Basket Analysis terhadap data Penjualan Produk Buku dengan menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth) Studi Kasus Percetakan PT Gramedia. *Telematika MKOM*, 118-132.
- Suryadi, S. S. (2010). Menggali Pengetahuan dan Bongkahan Data. In S. S. Suryadi, *Menggali Pengetahuan dan Bongkahan Data*. Yogyakarta: Andi Publisher.
- Triyanto, W. A. (2014). Association Rule Mining untuk Penentuan Rekomendasi Promosi Produk. *Simetris*, 121-126.
- Wiyana, S. K. (2018). Analisis Algoritma FP Growth untuk Rekomendasi Produk pada Data Retail Penjualan Produk Kosmetik (Studi Kasus: MT Shop Kelapa Gading). *Seminar Nasional Teknologi Informasi*, 61-69.