

PAPER • OPEN ACCESS

Product recommendation system on building materials shopping using FP-Growth algorithm

To cite this article: Bagus Mulyawan *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **1007** 012144

View the [article online](#) for updates and enhancements.



ECS **240th ECS Meeting**
Digital Meeting, Oct 10-14, 2021
We are going fully digital!
Attendees register for free!
REGISTER NOW

Product recommendation system on building materials shopping using FP-Growth algorithm

Bagus Mulyawan,^{1*} Vionelsy¹, and Tri Sutrisno¹

¹Faculty of Information Technology Universitas Tarumanagara Jakarta Indonesia 11440

*bagus@untar.ac.id

Abstract. The application program created is an e-commerce website based application program that is used to provide product choices for building materials using the FP-Growth algorithm. This algorithm is used to find frequent itemset (data that most often appears) on store transaction data from January 2019 to the most recent transaction. The use of FP-Tree which is used simultaneously with the FP-Growth algorithm only scans the database 2 times. The first scanning is done to read the previous store sales transaction data. The second scanning on the tree trace to form a pattern. After obtaining a number of linkages that are higher than the minimum support from the transaction data, product choices can be generated on the website. This application was created using the Hypertext Processor (PHP) programming language base which is supported by the MySQL database and SQL Server. It is expected that the presence of web-based applications can improve store sales and the marketing process of building products for the better.

1. Introduction

Along with increasing population growth, economic growth also increased. This causes many entrepreneurs in the field of building materials to compete to build a spending system that can provide convenience for individuals. The need for building materials for construction needs has become one of the conditions that must exist in a development. As for some of the obstacles encountered such as each individual must visit a place that sells building materials every time they want to make a product purchase or want to see the product provided. This is considered not very effective and efficient because of the amount of time that has been spent on the trip. In addition, the number of product orders made by the buyer is also directly related to the amount of stock of goods in the warehouse. The difficulty of the seller to calculate the amount of stock of goods available at the warehouse is also one of the obstacles for the seller. This is because the seller cannot predict buyer behavior. The use of the website for a means of selling products is more efficient. In the system that is to be built, the recording of the amount of stock of goods, information related to the product, and every transaction made is stored in a database. In addition, the system to be built can provide product recommendations to buyers to be purchased simultaneously with other products.

2. Method and materials

2.1. Sample data

Examples of transactions taken are previous store transactions. Transactions taken as many as 5 transactions. The sales transaction data can be seen in the following Table 1:



Table 1. Sales Transaction Data

No	Transaction	Itemset
1	T01	BES10, BES8, STR50, KTB0
2	T02	BES10, BES8, STR50, KTB0, HBL75, SM380
3	T03	SM380, HBL75, STR50, KTB0
4	T04	BES10, SJKT40
5	T05	STB0, SJKT40

Note : BES10(besi 10 mm), BES8(besi 8 mm), STR50(semen TR 50 kg), KTB0(kawat bendrat), HBL75(hebel 7.5 cm), SM380(sakmu 380), SJKT40(semen JJKT 40 kg), STB0(sarung tangan bitnik).

2.2. Method

The method to be applied is the FP-Growth algorithm. FP-Growth algorithm is one algorithm that can be used to determine the set of data that often appears (frequent itemset) in a data set. FP-Growth algorithm is the development of Apriori algorithm. So the deficiencies in the Apriori algorithm are corrected by the FP-Growth algorithm[1]. Apriori's algorithm requires generating candidates to get frequent itemset. However, the FP-Growth algorithm does not generate candidates because FP-Growth uses the concept of tree development in searching for frequent itemset that provides a faster calculation than the Apriori algorithm[2]. This method can be used to search for product information which can be purchased simultaneously in a transaction database of goods demand.[3] These three stages will be repeated repeatedly for each product in the header table sorted by frequency, namely[4]:

1. The generation phase of the conditional pattern base
The conditional pattern base contains the prefix path and the suffix pattern. The generation of a conditional pattern base can be obtained through a FP-Tree that has been built before.
2. FP-Tree conditional generation stage
At this stage, the support count of each item on the conditional pattern base is summed. For counts of support count greater than the minimum support limit will be raised on the conditional FP-Tree.
3. The frequent itemset search stage.
In searching for a frequent itemset, if the FP-Tree conditional is a single path, a frequent itemset will be obtained by combining items for each FP-Tree conditional. If it is not a single trajectory, then recursive generation of FP-Growth is carried out.

3. Results and discussion

The sales transaction data can be seen in the following Table 1. In addition, the number of items appearing is calculated. The occurrence value of an item that is greater than the minimum support count will be raised on the conditional pattern base. The results of frequent each item can be seen in Table 2.

Table 2. Fequet Item

No.	Item	Frequent item
1	BES10	3
2	KTB0	3
3	STR50	3

4	BES8	2
5	HBL75	2
6	SJKT40	2
7	SM380	2

After a frequent itemset is obtained, the next step is to create an FP-Tree based on the largest frequent items. If there is an itemset that has a different prefix, then the next itemset is built on a different path[5]. FP-Tree Development is divided into three steps, namely[6]:

1. Scanning data set to determine the amount of support for each item.
2. Remove unnecessary data and make a list order of the most items appearing in descending order.
3. Data scanning specifies one transaction at a time to create an FP-Tree. If a unique transaction will form a new path and a new node with a count starting from 1. The advantage of FP-Tree is that it only uses two data scanning.

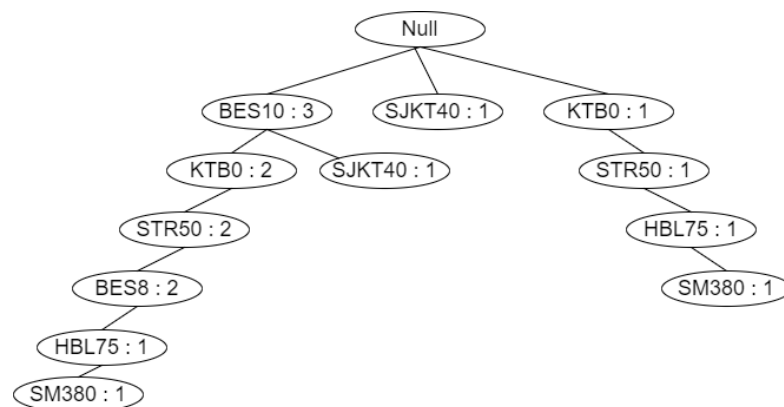


Figure 1. FP-Tree

After the formation of FP-Tree is carried out, it is continued by searching for a conditional pattern base containing the initial path (prefix path) and suffix pattern. When determining the conditional pattern base, what is considered is the suffix pattern. The search for an item for a conditional pattern base starts by looking at the number of occurrences and their relation to the suffix pattern. The support count of each item on each conditional pattern base is summed, after that every item that is greater than the minimum support will be raised with a conditional FP-Tree as shown in Table 3 below:

Table 3. The result of conditional pattern base and FP-Tree

<i>Item</i>	<i>Conditional Pattern Base</i>	<i>Conditional FP-Tree</i>
SM380	{BES10, KTBO, STR50, BES8, HBL75:1}, {KTBO, STR50, HBL75:1}	{KTBO:2}, {STR50:2}, {HBL75:2}
SJKT40	{BES10:1}	-
HBL75	{BES10, KTBO, STR50, BES8:1}, {KTBO, STR50:1}	{KTBO:2},{STR50:2}
BES8	{BES10, KTBO, SRT50:2}	{BES10:2}, {KTBO:2},{STR50:2}
STR50	{BES10, KTBO:2}, {KTBO:1}	{BES10:2, KTBO:3}

KTBO	{BES10:2}	{BES10:2}
BES10	-	-

The next step after getting a conditional FP-Tree is to look for frequent itemset. In frequent pattern generated, items will be combined items for each FP-Tree conditional. For frequent pattern generated results, see Table 4 below:

Table 4. Result of frequent pattern generated

Item	Frequent Pattern Generated
SM380	{KTBO, SM380: 2}, {STR50, SM380:2}, {HBL75, SM380:2}
SJKT40	-
HBL75	{KTBO, HBL75:2}, {STR50, HBL75:2}
BES8	{BES10, BES8:2}, {KTBO, BES8: 2}, {STR50, BES8:2}
STR50	{BES10, STR50:2},{KTBO,STR50:3}
KTBO	{BES10, KTBO:2}
BES10	-

To determine confidence, you can use the equation (1) as follows [2]:

$$Confidence(A \rightarrow B) = P(A|B) = \frac{\text{number of transactions containing A and B}}{\text{number of transactions containing A}} \quad (1)$$

The following is an example of calculating the confidence value by looking at Table 5 below:

Table 5. The results of confidence value calculation

No.	Rule	confidence
1.	BES8-> BES10	$\frac{\text{transaction}(\text{BES8 dan BES10})}{\text{transaction BES8}} = \frac{2}{3} = 66,66667 \%$
2.	BES8->KTBO	$\frac{\text{transaction}(\text{BES8 dan KTBO})}{\text{transaction BES8}} = \frac{2}{3} = 66,66667 \%$
3.	BES8->STR50	$\frac{\text{transaction}(\text{BES8 dan STR50})}{\text{transaction BES8}} = \frac{2}{3} = 66,66667 \%$
4.	HBL75->KTBO	$\frac{\text{transaction}(\text{HBL75 dan KTBO})}{\text{transaction HBL75}} = \frac{2}{3} = 66,66667 \%$
5.	HBL75->STR50	$\frac{\text{transaction}(\text{HBL75 dan STR50})}{\text{ransaction HBL75}} = \frac{2}{3} = 66,66667 \%$
6.	SM380->KTBO	$\frac{\text{transaction}(\text{SM380 dan KTBO})}{\text{transaction SM380}} = \frac{2}{3} = 66,66667 \%$
7.	SM380->STR50	$\frac{\text{transaction}(\text{SM380 dan STR50})}{\text{ransaction SM380}} = \frac{2}{3} = 66,66667 \%$
8.	SM380->HBL75	$\frac{\text{transaction}(\text{SM380 dan HBL75})}{\text{transaction SM380}} = \frac{2}{2} = 100 \%$
9.	KTBO->BES10	$\frac{\text{transaction}(\text{KTBO dan BES10})}{\text{transaction KTBO}} = \frac{2}{3} = 66,66667 \%$

10.	STR50->BES10	$= \frac{\text{transaction}(\text{STR50 dan BES10})}{\text{transaction STR50}} = \frac{2}{3} = 66,66667 \%$
11.	STR50->KTBO	$= \frac{\text{transaction}(\text{STR50 dan KTBO})}{\text{transaction STR50}} = \frac{2}{2} = 100 \%$

Based on the results of the calculation of confidence, the rules obtained can be seen in Table 6.

Table 6. Rules of items

No.	Rule	Confidence
1	If you buy BES8, you will get BES10 product recommendations	66,66667%
2	If you buy BES8 you will get KTBO product recommendations	66,66667%
3	If you buy BES8 you will get STR50 product recommendations	66,66667%
4	If you buy HBL75 you will get KTBO product recommendations	66,66667%
5	If you buy HBL75 you will get STR50 product recommendations	66,66667%
6	If you buy SM380 you will get KTBO product recommendations	66,66667%
7	If you buy SM380 you will get STR50 product recommendations	66,66667%
8	If you buy SM380 you will get HBL75 product recommendations	100%
9	If you buy KTBO you will get BES10 product recommendations	66.66667%
10	If you buy STR50 you will get BES10 product recommendations	66.66667%
11	If you buy STR50 you will get KTBO product recommendations	100%

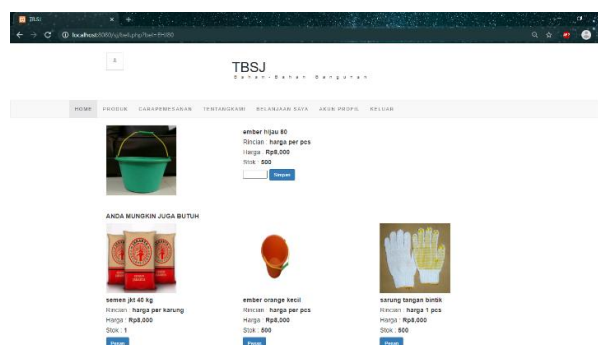


Figure 2. Web page display for product recommendations

4. Conclusion

Based on the results of tests conducted on the application program of the product selection recommendation system in the application of spending on building materials using the web-based FP-GROWTH algorithm, it can be concluded that:

1. This application program has been running well and can be used to facilitate the sales of SJ Building Stores sales without the need to go through face to face.

2. This application program is in accordance with the needs desired by the store.
3. This application program can produce product choices using the FP-Growth algorithm based on existing transaction data.

5. References

- [1] Kusrini and Luthfi, E. T. 2009 *Data Mining Algorithm* (Yogyakarta).
- [2] Erwin, 2009 Market Basketball Analysis with Apriori Algorithms and FP-Growth Generic Journal, Vol.4, 2-27.
- [3] Ikhwan, A., Nofriansyah, D. and Sriani 2015 *Application of Data Mining with the FP-Growth Algorithm to Support Educational Promotion Strategies* (Case Study of Triguna Dharma STMIK Campus) SAINTIKON Scientific Journal. **14** -211-226.
- [4] Han, J., Kamber, M., and Pei, J. 2006 *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers.
- [5] Chen, Ming-Chin, S.J. Cheng, and Y.Hwang. 2005 *An empirical investigation of the relationship between intellectual capital and firm's market value and financial performance*. *Journal of Intellectual Capital*, Vol.6, Issue 2.
- [6] Han, J., Kamber, M., and Pei, J. 2011 *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers.