

Implementation of FP-Growth and Fuzzy C-Covering Algorithm based on FP-Tree for Analysis of Consumer Purchasing Behavior

Rosya Rachmania
Gunadarma University
Depok, Indonesia

Raden Supriyanto
Gunadarma University
Depok, Indonesia

ABSTRACT

The FP-Growth and Fuzzy C-Covering algorithms are known to correct the Apriori weakness. FP-Growth uses the FP-Tree technique which is famous for the divide and conquer methods and does not generate itemset candidate generation. Fuzzy C-Covering uses the max item threshold technique to limit the execution of transactions. This algorithm requires a large memory and long execution time because of repeated data scans, since it is implemented to FP-Tree. The sales transaction data for IKK cooperative in 2018 amounted to 51,384 data. Data is used to identify items that might be purchased together with other items. Currently cooperatives do not have a data processing system for analysis of consumer buying patterns. Research is conducted to find association rules by implementing FP-Growth and Fuzzy C-Covering algorithms based on FP-Tree and to measure performance between algorithms based on execution time, memory usage, and the accuracy of association rules. Based on the test results, Fuzzy c-Covering based on FP-Tree uses less memory because the results of the tree formation are not stored and the execution time is longer because it is defined in the fuzzy set. FP-Growth has higher accuracy with the resulting association rules is risoles rahmat, tahu isi emly, pastel bihun susi with support 0.023%, and confidence 100%. Whereas Fuzzy c-Covering based on FP-Tree generates aqua 600ml, nasi telur balado siska, tahu bakso siska with support 0.05%, and confidence 21%.

General Terms

Data mining, Association Algorithm.

Keywords

Data Mining, Association Rule Mining, FP-Growth, Fuzzy c-Covering, FP-Tree.

1. INTRODUCTION

Association rule mining is a method of finding relationship patterns between data in a data set. Based on the pattern, the emergence of data can predict the appearance of other data (Adinugroho and Sari, 2018). The Frequent Pattern Growth (FP-Growth) algorithm and the Fuzzy C-Covering algorithm are known to correct the Apriori algorithm weaknesses. The FP-Growth algorithm searches for an item without generating candidate itemsets. The technique used by the FP-Growth algorithm is a Frequent Pattern Tree (FP-Tree) technique famous for the Divide and conquer methods, in the formation of the association rules this method works by dividing the item into smaller pieces and then Process the data and complete each section recursively.

The Fuzzy C-Covering algorithm is a generalization of the Fuzzy method C-Partition. This algorithm has the perception

that more and more items on a transaction then the smaller the level of association is generated. The Fuzzy C-Covering is very attentive to the items per transaction relationship, so each item is defined as a fuzzy set. This algorithm uses the Max item threshold technique by giving limits to not executing transactions over a predetermined threshold such as its perception. However this algorithm requires a large memory in search of an itemset combination because it performs the same repetitive data scan as the Apriori algorithm does. Therefore, it applies FP-Tree in the lookup combination itemset to reduce memory usage and excessive execution time.

Ikatan Kekeluargaan Karyawan (IKK) Cooperative is a consumer cooperative owned by Pondok Indah Hospital, established in the year 2014. This cooperative provides consumer goods with various types of food, beverages and other supplies to meet the needs of consumers who are inside and outside the hospital environment. The sales transaction data owned by IKK Cooperative in 2018 amounted to 51,384 data (cooperative IKK, 2018). The Data is used for consumer purchase behavior analysis by identifying items that are likely to be purchased alongside other items based on the set of item combinations and designing sales strategies such as product layout. But nowadays, the IKK cooperative does not have a system to analyze the sales data. Therefore, it takes the association data mining method to analyze large amounts of data to find the frequency of the items that have a relationship pattern. Based on this, then researched analysis consumer purchase behavior on the cooperative IKK by processing data using systems that implement the FP-Growth algorithm and the Fuzzy algorithm C-Covering-based FP-Tree and determine the algorithm Which is more accurate and efficient which can be used as a recommendation in data mining.

2. LITERATURE REVIEW

2.1 Association Rule Mining

Association rule mining is a method of finding relationship patterns between data in a data set. Based on the pattern, the emergence of data can predict the appearance of other data. The Association rule aims to find relationships between the data in a large set of data (Adinugroho and Sari, 2018).

The Association rule will find a specific pattern that associates one data with other data. Then to look for the association rule of a data set, the first step to be done is to look for the frequency of itemset first. After all the frequent itemset patterns are found, then look for associative rules or qualified rules of association that have been determined.

Analysis of association or association rule mining is a data mining technique to find the rules of association between a combination of items. An example of an association of

purchase analysis in a supermarket is that it can know how likely a customer is to buy bread along with milk. With such knowledge the supermarket owners can arrange the placement of the goods or design a marketing campaign by using discount coupons for a certain combination of items. The analysis of the association became famous for its application to analyze shopping cart contents in supermarkets. The analysis of the association is also known as one of the data mining techniques that become the basis of various other data mining techniques (Vulandari, 2017).

2.2 FP-Growth Algorithm

According to Goldie (2012), Frequent Pattern Growth (FP-Growth) is an alternative algorithm that can be used to determine a frequent itemset of data set in a dataset (Sepri and Afdal, 2017). The characteristics of the FP-Growth algorithm use the data tree structure called FP-Tree. By using FP-Tree, the FP-Growth algorithm can directly extract the frequent Itemset from FP-Tree. Extracting Itemset that has a high frequency using the FP-Growth algorithm will be carried out by generating the structure of the data tree (Astrina, Arifin, and Pujianto, 2019).

The FP-Growth algorithm generates a frequent Itemset of FP-Tree using the Divide and Conquer method. To find a frequent Itemset without creating an item candidate generation, built using 2 datasets, in the path 1 scans the transaction database and finds the support value for each item, then gradually the support value will increase and items that have frequencies below minimum support will be wasted, then sort the support values based on the highest frequency. In Path 2, FP-Growth reads transactions at the same time and tracks them to the path (Mayilvaganan and Kalpanadevi, 2018).

This algorithm seeks an Itemset without doing a generation candidate Itemset. The FP-Growth method can be divided into 3 main phases (Purba and Siswanto, 2015):

1. The generation phase of the conditional pattern base.
The Conditional pattern base is a sub-database containing the prefix path and suffix pattern. The generation of conditional pattern base was obtained through the previously built FP-Tree.
2. The FP-Tree conditional generation stage.
At this stage, the support count of each item in each conditional pattern base is added, then each item that has the number of support count is greater than or equal to the minimum support count to be raised with FP-Tree conditionals.
3. Search stage of frequent Itemset.
If the FP-Tree conditional is a single path, a frequent pattern is obtained by combining items for each FP-Tree conditional. If it is not a single path, the FP-Growth generation is recursively.

2.3 Fuzzy c-Covering Algorithm

The Fuzzy C-Covering is one of the methods used to classify the elements of a universal set into partitions of fuzzy sets. The fuzzy C-Covering itself is a generalization of the Fuzzy method C-Partition which is known to fix the weakness of the Apriori algorithm. This method has the perception that more items are purchased in one transaction then the relationship between items is weaker. In practice, fuzzy conditional probability relation can be used as the basis for representing a degree of similarity relationship between the two fuzzy sets in Universe U. In the fuzzy definition of conditional probability

relation, the value Probability can be estimated based on the semantic relationship between fuzzy sets by using the subjective view of the probability theory (Budhi, Lim, and Prayitno, 2005).

Here are the steps of the Fuzzy c-Covering algorithm (Angraini, Indwiarti, and Nhita, 2018):

1. Determine the max item threshold required. Max item threshold is a delimiter used to filter transactions based on the number of items in the transaction.
2. Look for the records in the transaction table that satisfy Max item threshold and save it into QT, where:

$$QT = \{t \mid |t| \leq \text{ith}, \text{ith} \in \text{positive integer}\}.$$

Where:

QT (Qualified Transaction): The set of transactions that meet the max item threshold; t: transactions; | t |: The number of items in a transaction. ith: Max item threshold.

3. Set k = 1 (k is the variable to specify the number of combinations).
4. Determine the min_support to-K.
5. Looking for support of any combination of K-items that allow existing in the transaction with the formula:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

6. Conducting a supervisor against the combination of items that are in the transaction that does not meet.
7. Set k = k + 1, where if k > ith.
8. Defines each item that has been obtained from the above steps as a fuzzy set (called the fuzzy Set item) against the QT transaction.
9. Find the candidate rules by calculating the confidence of each K-item combination that satisfies the K min_support of the fuzzy Itemset with the formula:

$$\text{Confidence } P(A|B) = \frac{\text{Support (A} \cap \text{B)}}{P(A)}$$

Table 1. Related studies

Year	Author's name	Title	Method	Result
2017	Sinthuja, M., Puviarasan, N., and Aruna, P	Evaluating the Performance of Association Rule Mining Algorithms	Apriori, ECLAT, FP-Growth Algorithm	The resulting analysis indicates that the runtime and memory algorithms differ for different datasets. After evaluating the results of the experiment based on performance characteristics

				such as runtime and memory usage, it is demonstrated that the performance of the FP-Growth algorithm is better than the Apriori and ECLAT algorithms.
2016	Bala, A., Shuaibu, M. Z., Lawal, Z. K., and Zakari, R. Y	Performance Analysis of Apriori and FP-Growth Algorithms (Association Rule Mining)	Apriori and FP-Growth Algorithm	The FP-Growth algorithm is faster in terms of execution time compared to the Apriori algorithm. This indicates that the amount of time required to run up to completion is less than the amount of time required by the Apriori algorithm.
2012	Rindengan, A. J	Comparative Association Rule of binary-shaped and Fuzzy C-Partition in the Basket Market analysis in Data Mining	Association Rule of binary-shaped and Fuzzy C-Partition	<p>This research discusses the association's rules by considering the number of items purchased in one transaction. The assumption is that the relation of buying an item with another item in one transaction will be smaller, if the number of items purchased is more and more.</p> <p>Based on the results of the analysis the transaction table can be expressed in the form of fuzzy sets (with the assumption). Compared to using the binary-shaped association rule, the fuzzy C-partition gives support and</p>

				confidence values tend to be smaller but provides more realistic results.
2018	Angraini, K. N., Indwiarti, and Nhita, F	Implementation of Fuzzy c-Covering Algorithm to Identify Purchase Patterns in Supermarket Transaction Data.	Fuzzy c-Covering Algorithm	<p>Fuzzy c-Covering method can be used to classify the elements of a universal set of overall product items in a minimarket into more focused and detailed partitions based on existing product items. Therefore, Fuzzy c-Covering is applied to overcome the obstacles that have occurred in the process of finding relationships between items. The analysis shows that the higher the maximum item threshold used, the higher the support value generated, but does not affect the confidence value generated and the time required is faster.</p>

Based on the related studies, the FP-Growth algorithm and the Fuzzy C-Covering algorithm have the same performance characteristics as speed in forming the relationship patterns between items compared to the Apriori algorithm. However in the case of memory usage of the Fuzzy algorithm C-Covering requires a large memory when performing a combination lookup of itemset because it is performed repeatedly scan data. Therefore, it applies FP-Tree in the lookup combination itemset to reduce memory usage and excessive execution time.

3. METHODOLOGY

3.1 Data Selection

Source of the research data obtained from the sales transaction data of IKK cooperative Pondok Indah Hospital in 2018. The attributes used are invoice number, item code, and item name.

Table 2. Attributes Selection

Attributes	Data Used
Date	No
No. Invoice	Yes
Item code	Yes
Item Name	Yes
Sale price	No
Cost of Sale	No
Discount	No
Profit or loss	No

3.2 Pre-processing

Before the process of data mining can be carried out, need to do the process of cleaning the data that is the focus of KDD. The cleaning process among other things removes data duplication, checks inconsistent data, and corrects errors in the data, such as typography errors.

3.3 Transformation

The sales transaction data of IKK cooperative PDF file is then transformed manually by converting the file format to CSV against selected attributes of the transaction data.

Table 3. Sales Transaction Data

No. Invoice	Item code	Item Name
R43-030118001	089686060027	POP MIE AYAM 75GR
R43-030118002	8995899250228	BONCABE RASA TERI 22.5G
R43-030118003	8992761122331	FRESSTE APEL 500ML

3.4 Data Mining

3.4.1 FP-Growth Algorithm

Here are the stages in the process of looking for product association rules using the FP-Growth algorithm:

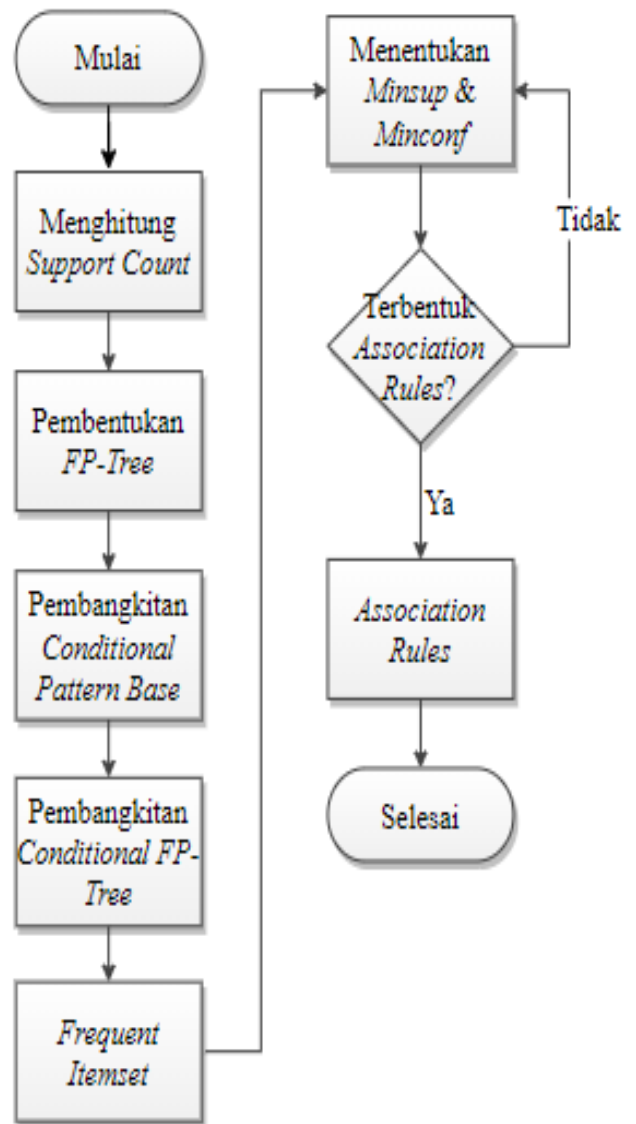


Figure 1. Flowchart FP-Growth algorithm

3.4.2 Fuzzy c-Covering Algorithm based on FP-Tree

Here are the stages in the process of looking for product association rules using the Fuzzy C-Covering algorithm based on FP-Tree:

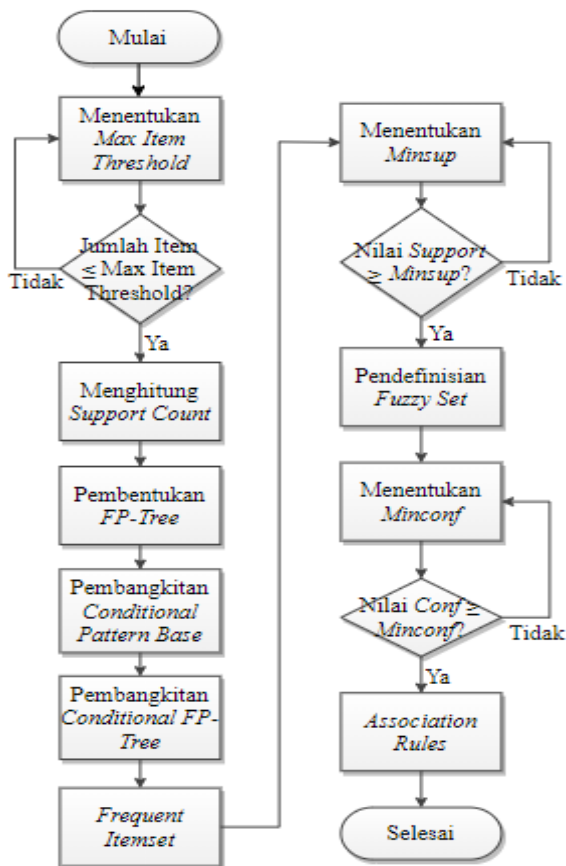


Figure 2. Flowchart Fuzzy c-Covering algorithm based on FP-Tree

4. RESULT AND DISCUSSION

4.1 Implementation of FP-Growth Algorithm

4.1.1 Calculating the Support Count

Calculates the support count by looking at the frequency of items appearing on a transaction. A minimum support $\geq 0,1$ (support count ≥ 2) is determined, calculating the value of support for each item based on transaction data in Table 5 in the following way:

$$\text{Support (A)} = 9/20 = 0,45$$

$$\text{Support (B)} = 9/20 = 0,45$$

$$\text{Support (C)} = 4/20 = 0,2$$

$$\text{Support (D)} = 4/20 = 0,2$$

$$\text{Support (E)} = 4/20 = 0,2$$

Table 4. Support count and Support value item

Item Name	Support Count	Support Value
Arem-Arem Fitri (A)	9	0,45
Bakwan Jagung Emly (B)	9	0,45
Sosis Solo (I)	8	0,4
Sus Rahmat (J)	8	0,4
Pastel Ayam Kus (F)	7	0,35

Piscok Meleleh (G)	6	0,3
Sego Kucing Hesti (H)	6	0,3
Ketupat Sayur Dian (C)	4	0,2
Martabak Daging Rahmat (D)	4	0,2
Nasi Ayam Sambal Ijo Emly (E)	4	0,2

Table 5. Sample data based on biggest support count

No. Invoice	Item Name
R43-010818001	Bakwan Jagung Emly (B), Sus Rahmat (J), Nasi Ayam Sambal Ijo Emly (E)
R43-010818002	Arem-Arem Fitri (A), Pastel Ayam Kus (F), Nasi Ayam Sambal Ijo Emly (E)
R43-010818003	Sosis Solo (I), Pastel Ayam Kus (F)
R43-010818004	Bakwan Jagung Emly (B), Piscok Meleleh (G)
R43-010818005	Arem-Arem Fitri (A), Sosis Solo (I), Sus Rahmat (J)
R43-010818006	Sosis Solo (I), Piscok Meleleh (G)
R43-010818007	Bakwan Jagung Emly (B), Piscok Meleleh (G)
R43-010818008	Arem-Arem Fitri (A), Sosis Solo (I), Sego Kucing Hesti (H)
R43-010818009	Arem-Arem Fitri (A), Sosis Solo (I)
R43-010818010	Bakwan Jagung Emly (B), Sego Kucing Hesti (H), Martabak Daging Rahmat (D)
R43-010818011	Bakwan Jagung Emly (B), Sus Rahmat (J), Martabak Daging Rahmat (D)
R43-010818012	Arem-Arem Fitri (A), Pastel Ayam Kus (F), Ketupat Sayur Dian (C)
R43-010818013	Sosis Solo (I), Pastel Ayam Kus (F)
R43-010818014	Pastel Ayam Kus (F), Ketupat Sayur Dian (C)
R43-010818015	Arem-Arem Fitri (A), Bakwan Jagung Emly (B), Sus Rahmat (J), Pastel Ayam Kus (F), Martabak Daging Rahmat (D)
R43-010818016	Bakwan Jagung Emly (B), Sus Rahmat (J), Sego Kucing Hesti (H), Ketupat Sayur Dian (C), Martabak Daging Rahmat (D)
R43-010818017	Arem-Arem Fitri (A), Pastel Ayam Kus (F), Piscok Meleleh (G), Ketupat Sayur Dian (C), Nasi Ayam Sambal Ijo Emly (E)
R43-010818018	Bakwan Jagung Emly (B), Sus Rahmat (J), Piscok Meleleh (G), Sego Kucing Hesti (H), Nasi Ayam Sambal Ijo Emly (E)

R43-010818019	Arem-Arem Fitri (A), Bakwan Jagung Emly (B), Sosis Solo (I), Sus Rahmat (J), Sego Kucing Hesti (H)
R43-010818020	Arem-Arem Fitri (A), Sosis Solo (I), Sus Rahmat (J), Piscok Meleleh (G), Sego Kucing Hesti (H)

4.1.2 FP-Tree Formation

FP-Tree formation uses the data contained in Table 5.

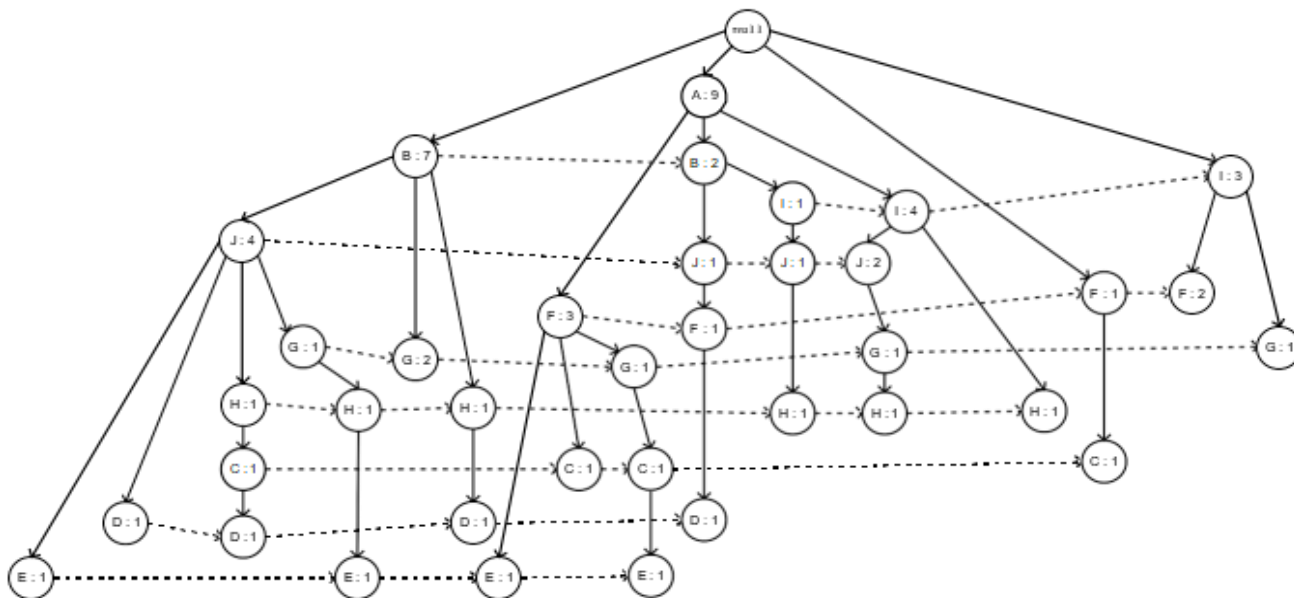


Figure 3. FP-Tree formation after reading Invoice R43-010818020

4.1.3 Conditional Pattern Base Generation

A conditional pattern base generation by separating all path ends based on an item that has the smallest support count. The following conditional pattern base generation processes against the path ending with vertex E:

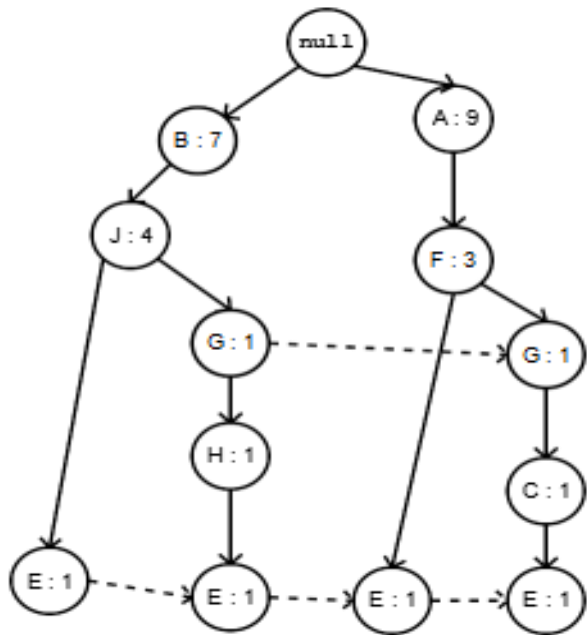


Figure 4. Paths Containing Node E

The following data items are obtained after the conditional generation pattern base:

Table 6. Conditional pattern base

Item	Conditional Pattern Base
E	{J, B : 1}, {H, G, J, B : 1}, {F, A : 1}, {C, G, F, A : 1}
D	{J, B : 1}, {C, H, J, B : 1}, {H, B : 1}, {F, J, B, A : 1}
C	{H, J, B : 1}, {F, A : 1}, {G, F, A : 1}, {F : 1}
H	{J, B : 1}, {G, J, B : 1}, {B : 1}, {J, I, B, A : 1}, {G, J, I, A : 1}, {I, A : 1}
G	{J, B : 1}, {B : 2}, {F, A : 1}, {J, I, A : 1}, {I : 1}
F	{A : 3}, {J, B, A : 1}, {I : 2}
J	{B : 4}, {B, A : 1}, {I, B, A : 1}, {I, A : 2}
I	{B, A : 1}, {A : 4}, {I : 3}
B	{A : 2}

4.1.4 Conditional FP-Tree Generation

After generating a conditional pattern base on a path ending in vertex E, FP-Tree will be generated by removing a vertex that ends with E due to a support count belonging to vertex E = 1. Then recalculate the support count above the vertex E corresponds to the frequency of the vertex emergence E. Change vertex E after generating a conditional FP-Tree:

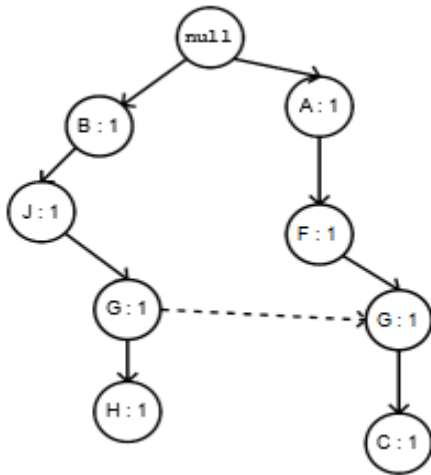


Figure 5. Conditional FP-Tree for node E

Perform FP-Tree conditional generation against other nodes such as Vertex E. The following FP-Tree conditional generation result of the item:

Table 7. Conditional FP-Tree

Item	Conditional FP-Tree
E	{J, B : 2}, {G : 2}, {F, A : 2}
D	{J, B : 3}, {H, B : 2}
C	{F, A : 2}
H	{J, B : 3}, {G : 2}, {I, A : 3}
G	{B : 2}, {A : 2}, {I : 2}, {J : 2}
F	{A : 4}, {I : 2}
J	{B : 4}, {B, A : 2}, {I, A : 2}
I	{A : 4}, {I : 3}
B	{A : 2}

4.1.5 Frequent Itemset

After FP-Tree conditional generation acquired the frequent Itemset as follows:

Table 8. Frequent Itemset

Item	Frequent Itemset
E	{E, J, B : 2}, {E, G : 2}, {E, F, A : 2}
D	{D, J, B : 3}, {D, H, B : 2}
C	{C, F, A : 2}
H	{H, J, B : 3}, {H, G : 2}, {H, I, A : 3}
G	{G, B : 2}, {G, A : 2}, {G, I : 2}, {G, J : 2}
F	{F, A : 4}, {F, I : 2}
J	{J, B : 4}, {J, B, A : 2}, {J, I, A : 2}
I	{I, A : 4}, {I : 3}
B	{B, A : 2}
A	{A : 9}

The Frequent Itemset to be used in the search association rules is {E, J, B}, {E, G}, {E, F, A}, {D, J, B}, {D, H, B}, {C, F, A}, {H, J, B}, {H, G}, {H, I, A}, {G, B}, {G, A}, {G, I}, {G, J}, {F, A}, {F, I}, {J, B}, {J, B, A}, {J, I, A}, {I, A}, {B, A}.

4.1.6 Looking for Association Rules

Look at the product association rules by performing confidence value calculations against the frequent Itemset. The following are calculation of confidence values for frequent Itemsets:

$$\text{confidence}(E, J \rightarrow B) = \frac{2}{2} \times 100 = 100\%$$

$$\text{confidence}(E, B \rightarrow J) = \frac{2}{2} \times 100 = 100\%$$

$$\text{confidence}(J, B \rightarrow E) = \frac{2}{6} \times 100 = 33,3\%$$

$$\text{confidence}(E \rightarrow J, B) = \frac{2}{4} \times 100 = 50\%$$

$$\text{confidence}(J \rightarrow E, B) = \frac{2}{8} \times 100 = 25\%$$

$$\text{confidence}(B \rightarrow E, J) = \frac{2}{9} \times 100 = 22,2\%$$

$$\text{confidence}(E \rightarrow G) = \frac{2}{4} \times 100 = 50\%$$

$$\text{confidence}(G \rightarrow E) = \frac{2}{6} \times 100 = 33,3\%$$

Calculate confidence values for all the frequent itemsets as in Itemset {E, J, B} and {E, G}.

Confidence value $\geq 90\%$, following the eligible association rules:

- If buy Nasi Ayam Sambal Ijo Emly (E) and Sus Rahmat (J) then will buy Bakwan Jagung Emly (B) with a confidence value of 100%.
- If buy Nasi Ayam Sambal Ijo Emly (E) and Bakwan Jagung Emly (B) then will buy Sus Rahmat (J) with a confidence value of 100%.
- If buy Nasi Ayam Sambal Ijo Emly (E) and Pastel Ayam Kus (F) then will buy Arem-Arem Fitri (A) with a confidence value of 100%.
- If buy Nasi Ayam Sambal Ijo Emly (E) and Arem-Arem Fitri (A) then will buy Pastel Ayam Kus (F) with a confidence value of 100%.
- If buy Martabak Daging Rahmat (D) and Sus Rahmat (J) then will buy Bakwan Jagung Emly (B) with a confidence value of 100%.
- If buy Martabak Daging Rahmat (D) and Sego Kucing Hesti (H) then will buy Bakwan Jagung Emly (B) with a confidence value of 100%.
- If buy Ketupat Sayur Dian (C) and Arem-Arem Fitri (A) then will buy Pastel Ayam Kus (F) with a confidence value of 100%.
- If buy Sego Kucing Hesti (H) and Sosis Solo (I) then will buy Arem-Arem Fitri (A) with a confidence value of 100%.
- If buy Sego Kucing Hesti (H) and Arem-Arem Fitri (A) then will buy Sosis Solo (I) with a confidence value of

100%.

- If buy Bakwan Jagung Emly (B) and Arem-Arem Fitri (A) then will buy Sus Rahmat (J) with a confidence value of 100%.

4.2 Implementation of Fuzzy c-Covering Algorithm Based on FP-Tree

4.2.1 Determine the Max Item Threshold

Determined Max item threshold ≤ 4 then transactions that have several items greater than 4 will be ignored by the system.

4.2.2 Calculating the Support Count

Calculates the support count by looking at the frequency of occurrence of items based on the transaction data in Table 10. Here is a support count for each item:

Table 9. Support count

Item Name	Support Count
Sosis Solo (I)	6
Arem-Arem Fitri (A)	5
Bakwan Jagung Emly (B)	5
Pastel Ayam Kus (F)	5
Piscok Meleleh (G)	3
Sus Rahmat (J)	3
Ketupat Sayur Dian (C)	2
Martabak Daging Rahmat (D)	2
Nasi Ayam Sambal Ijo Emly (E)	2
Sego Kucing Hesti (H)	2

	Rahmat (J), Nasi Ayam Sambal Ijo Emly (E)
R43-010818002	Arem-Arem Fitri (A), Pastel Ayam Kus (F), Nasi Ayam Sambal Ijo Emly (E)
R43-010818003	Sosis Solo (I), Pastel Ayam Kus (F)
R43-010818004	Bakwan Jagung Emly (B), Piscok Meleleh (G)
R43-010818005	Sosis Solo (I), Arem-Arem Fitri (A), Sus Rahmat (J)
R43-010818006	Sosis Solo (I), Piscok Meleleh (G)
R43-010818007	Bakwan Jagung Emly (B), Piscok Meleleh (G)
R43-010818008	Sosis Solo (I), Arem-Arem Fitri (A), Sego Kucing Hesti (H)
R43-010818009	Sosis Solo (I), Arem-Arem Fitri (A)
R43-010818010	Bakwan Jagung Emly (B), Martabak Daging Rahmat (D), Sego Kucing Hesti (H)
R43-010818011	Bakwan Jagung Emly (B), Sus Rahmat (J), Martabak Daging Rahmat (D)
R43-010818012	Arem-Arem Fitri (A), Pastel Ayam Kus (F), Ketupat Sayur Dian (C)
R43-010818013	Sosis Solo (I), Pastel Ayam Kus (F)
R43-010818014	Pastel Ayam Kus (F), Ketupat Sayur Dian (C)

Table 20. Sample Data based on Max Item Threshold and Support Count

No. Faktur	Nama Barang
R43-010818001	Bakwan Jagung Emly (B), Sus

4.2.3 FP-Tree Formation

FP-Tree formation uses the data contained in Table 10.

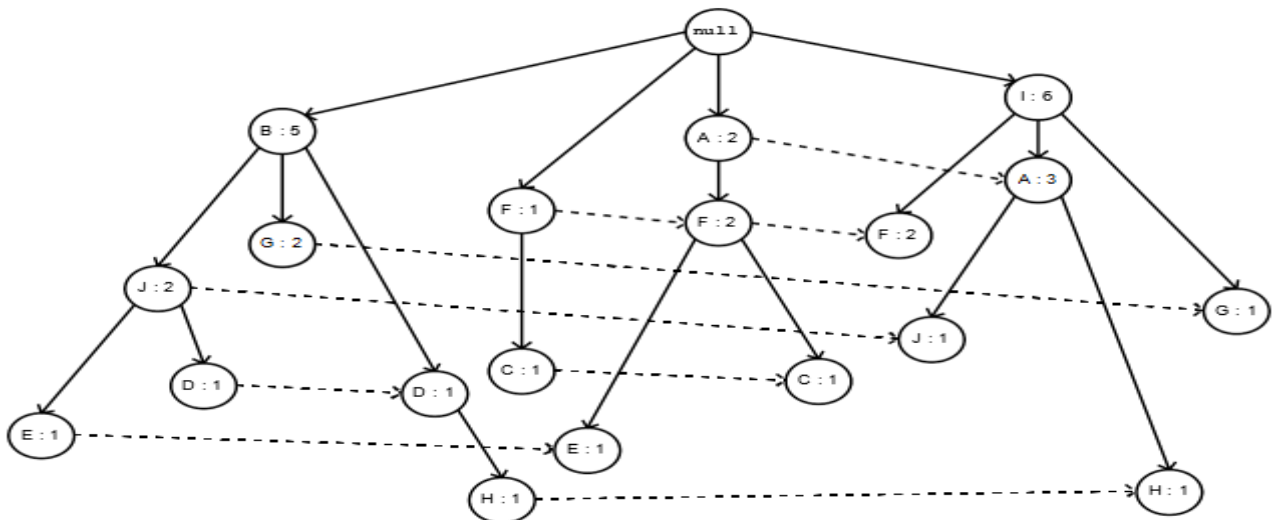


Figure 6. FP-Tree formation after reading Invoice R43-010818014

4.2.4 Conditional Pattern Base Generation

Table 11. Conditional Pattern Base

Item	Conditional Pattern Base
H	{D, B : 1}, {A, I : 1}
E	{J, B : 1}, {F, A : 1}
D	{J, B : 1}, {B : 1}
C	{F : 1}, {F, A : 1}
J	{B : 2}, {A, I : 1}
G	{B : 2}, {I : 1}
F	{A : 2}, {I : 2}
A	{I : 3}

4.2.5 Conditional FP-Tree Generation

Table 32. Conditional FP-Tree

Item	Conditional FP-Tree
D	{B : 2}
C	{F : 2}
J	{B : 2}
G	{B : 2}
F	{A : 2}, {I : 2}
A	{I : 3}

4.2.6 Frequent Itemset

Table 13. Frequent Itemset

Item	Frequent Itemset
D	{D, B : 2}
C	{C, F : 2}
J	{J, B : 2}
G	{G, B : 2}
F	{F, A : 2}, {F, I : 2}
B	{B : 5}
A	{A, I : 3}
I	{I : 6}

The Frequent itemset to be used in the search association rules is {D, B}, {C, F}, {J, B}, {G, B}, {F, A}, {F, I}, {A, I}.

4.2.7 Calculating Support values

Specified minimum support ≥ 0.05 (support count ≥ 2), calculates the support value $k = 1$, based on the transaction data in Table 10 in the following way:

$$\{A\} = \frac{0 + \frac{1}{3} + 0 + 0 + \frac{1}{3} + 0 + 0 + \frac{1}{3} + \frac{1}{2} + 0 + 0 + \frac{1}{3} + 0 + 0}{14} = 0,13$$

$$\{B\} = \frac{\frac{1}{3} + 0 + 0 + \frac{1}{2} + 0 + 0 + \frac{1}{2} + 0 + 0 + \frac{1}{3} + \frac{1}{3} + 0 + 0 + 0}{14} = 0,14$$

$$\{C\} = \frac{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \frac{1}{3} + 0 + \frac{1}{2}}{14} = 0,06$$

$$\{D\} = \frac{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \frac{1}{3} + \frac{1}{3} + 0 + 0 + 0}{14} = 0,05$$

$$\{E\} = \frac{\frac{1}{3} + \frac{1}{3} + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0}{14} = 0,05$$

$$\{F\} = \frac{0 + \frac{1}{3} + \frac{1}{2} + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}}{14} = 0,15$$

$$\{G\} = \frac{0 + 0 + 0 + \frac{1}{2} + 0 + \frac{1}{2} + \frac{1}{2} + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0}{14} = 0,11$$

$$\{H\} = \frac{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + \frac{1}{3} + 0 + \frac{1}{3} + 0 + 0 + 0 + 0}{14} = 0,05$$

$$\{I\} = \frac{0 + 0 + \frac{1}{2} + 0 + \frac{1}{3} + \frac{1}{2} + 0 + \frac{1}{3} + \frac{1}{2} + 0 + 0 + 0 + \frac{1}{2} + 0 + 0}{14} = 0,19$$

$$\{J\} = \frac{\frac{1}{3} + 0 + 0 + 0 + \frac{1}{3} + 0 + 0 + 0 + 0 + 0 + \frac{1}{3} + 0 + 0 + 0}{14} = 0,07$$

Calculate support values until no more items can be combined based on minimum support.

4.2.8 Defining Fuzzy Set

Each item that meets the minimum support is defined as a fuzzy set against the transaction. Here is defining item A into the fuzzy set against no invoice R43-010818002, R43-010818005, R43-010818008, R43-010818009 and R43-010818012:

$$\mu_A(R43 - 010818002) = \frac{\mu_{R43-002}(A)}{\mu_{R43-002}(A) + \mu_{R43-002}(F) + \mu_{R43-002}(E)}$$

$$= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{\frac{1}{3}}{1} = \frac{1}{3}$$

$$\mu_A(R43 - 010818005) = \frac{\mu_{R43-005}(A)}{\mu_{R43-005}(I) + \mu_{R43-005}(A) + \mu_{R43-005}(J)}$$

$$= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{\frac{1}{3}}{1} = \frac{1}{3}$$

$$\mu_A(R43 - 010818008) = \frac{\mu_{R43-008}(A)}{\mu_{R43-008}(I) + \mu_{R43-008}(A) + \mu_{R43-008}(H)}$$

$$= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{\frac{1}{3}}{1} = \frac{1}{3}$$

$$\mu_A(R43 - 010818009) = \frac{\mu_{R43-009}(A)}{\mu_{R43-009}(I) + \mu_{R43-009}(A)}$$

$$= \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{2}} = \frac{\frac{1}{2}}{1} = \frac{1}{2}$$

$$\mu_A(R43 - 010818012) = \frac{\mu_{R43-012}(A)}{\mu_{R43-012}(A) + \mu_{R43-012}(F) + \mu_{R43-012}(C)}$$

$$= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} = \frac{\frac{1}{3}}{1} = \frac{1}{3}$$

Define all items that meet the minimum support into the fuzzy set according to the no invoice contained in the item. Based on such calculations, it can be defined as follows:

$$\mu_A = \{(1/3)/002, (1/3)/005, (1/3)/008, (1/2)/009, (1/3)/012\}$$

$$\mu_B = \{(1/3)/001, (1/2)/004, (1/2)/007, (1/3)/010, (1/3)/011\}$$

$$\mu_C = \{(1/3)/012, (1/2)/014\}$$

$$\mu_D = \{(1/3)/010, (1/3)/011\}$$

$$\mu_E = \{(1/3)/001, (1/3)/002\}$$

$$\mu_F = \{(1/3)/002, (1/2)/003, (1/3)/012, (1/2)/013, (1/2)/014\}$$

$$\mu_G = \{(1/2)/004, (1/2)/006, (1/2)/007\}$$

$$\mu_H = \{(1/3)/008, (1/3)/010\}$$

$$\mu_I = \{(1/2)/003, (1/3)/005, (1/2)/006, (1/3)/008, (1/2)/009, (1/2)/013\}$$

$$\mu_J = \{(1/3)/001, (1/3)/005, (1/3)/011\}$$

4.2.9 Looking for Association Rules

Search the product association rules by performing confidence value calculations against the frequent itemset. The following are calculation of confidence values for frequent itemsets:

$$\text{confidence}(D \rightarrow B) = \frac{\frac{1}{3} + \frac{1}{3}}{\frac{1}{3} + \frac{1}{3}} \times 100 = 100\%$$

$$\text{confidence}(B \rightarrow D) = \frac{\frac{1}{3} + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3}} \times 100 = 33\%$$

$$\text{confidence}(C \rightarrow F) = \frac{\frac{1}{3} + \frac{1}{2}}{\frac{1}{3} + \frac{1}{2}} \times 100 = 100\%$$

$$\text{confidence}(F \rightarrow C) = \frac{\frac{1}{3} + \frac{1}{2}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \times 100 = 38,4\%$$

$$\text{confidence}(J \rightarrow B) = \frac{\frac{1}{3} + \frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} \times 100 = 66\%$$

$$\text{confidence}(B \rightarrow J) = \frac{\frac{1}{3} + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3}} \times 100 = 33\%$$

$$\text{confidence}(G \rightarrow B) = \frac{\frac{1}{2} + \frac{1}{2}}{\frac{1}{2} + \frac{1}{2} + \frac{1}{2}} \times 100 = 66,7\%$$

$$\text{confidence}(B \rightarrow G) = \frac{\frac{1}{2} + \frac{1}{2}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3}} \times 100 = 50\%$$

$$\text{confidence}(F \rightarrow A) = \frac{\frac{1}{3} + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \times 100 = 30,5\%$$

$$\text{confidence}(A \rightarrow F) = \frac{\frac{1}{3} + \frac{1}{3}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3}} \times 100 = 36\%$$

$$\text{confidence}(F \rightarrow I) = \frac{\frac{1}{2} + \frac{1}{2}}{\frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \times 100 = 46,3\%$$

$$\text{confidence}(I \rightarrow F) = \frac{\frac{1}{2} + \frac{1}{2}}{\frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \times 100 = 37,5\%$$

$$\text{confidence}(A \rightarrow I) = \frac{\frac{1}{3} + \frac{1}{3} + \frac{1}{2}}{\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3}} \times 100 = 63,4\%$$

$$\text{confidence}(I \rightarrow A) = \frac{\frac{1}{3} + \frac{1}{3} + \frac{1}{2}}{\frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}} \times 100 = 43,6\%$$

The selected association rules with confidence value $\geq 90\%$, are:

- If buy Martabak Daging Rahmat (D) then will buy Bakwan Jagung Emly (B) with a confidence value of 100%.
- If buy Ketupat Sayur Dian (C) then will buy Pastel Ayam Kus (F) with a confidence value of 100%.

4.3 Testing Data against Algorithms

The final phase is conducted testing of the IKK cooperative Sales transaction data using the FP-Growth algorithm and the Fuzzy C-Covering algorithm based on FP-Tree with a total of 51,384 transaction data. This test is conducted to determine performance between algorithms by comparing memory usage, execution time, and accuracy of product association rules.

4.3.1 Memory Usage (Bytes) Comparison

The following chart compares the memory usage size between the FP-Growth algorithm and the Fuzzy C-Covering algorithm based on FP-Tree.

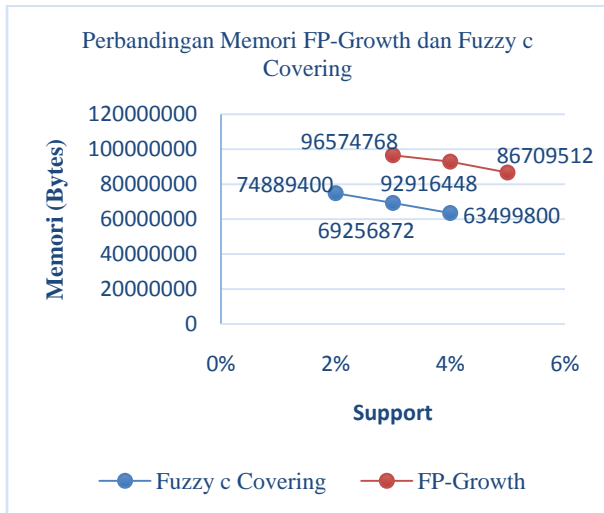


Figure 7. Memory Usage Comparison

4.3.2 Execution Time (s) Comparison

Figure 8 is a comparison chart of the execution time graph between the FP-Growth algorithm and the Fuzzy c-Covering algorithm based on FP-Tree in finding association rules.

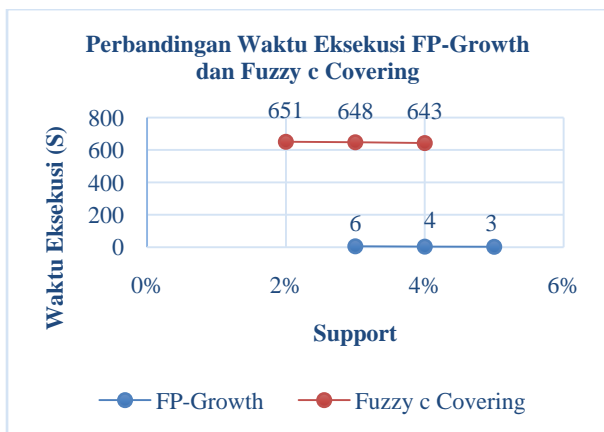


Figure 8. Execution Time Comparison

4.3.3 Accuracy of the Association Rules Between Algorithms

The accuracy of the association rules between the FP-Growth algorithm and the Fuzzy c-Covering algorithm based on FP-Tree is tested by comparing the confidence values of the association rules generated by each algorithm. The rules of product association produced by the FP-Growth algorithm and the Fuzzy c-Covering algorithm based on FP-Tree is based on the highest support and confidence values are summarized in Table 14.

Table 44. Association Rules of FP-Growth and Fuzzy c-Covering based on FP-Tree

Algorithm	Association Rules	Support value	Confidence value
FP-Growth	Risoles Rahmat, Tahu Isi Emly, Pastel Bihun Susi	0,023%	100%
Fuzzy c-Covering	Aqua 600ml, Nasi Telor Balado Siska, Tahu Bakso Siska	0,05%	21%

5. CONCLUSIONS AND SUGGESTIONS

5.1 Conclusions

1. The rules of the Product association on the data of the sales Transaction IKK Cooperative Pondok Indah Hospital were obtained by implementing the FP-Growth algorithm and the Fuzzy c-Covering algorithm based on FP-Tree.
2. The larger the minimum support for the FP-Growth algorithm and the Fuzzy c-Covering algorithm based on FP-Tree the smaller the memory is used. However, the memory used in the Fuzzy c-Covering algorithm based on FP-Tree is larger because this algorithm does not store Tree formation results during the search process of an itemset. The execution time of the Fuzzy c-Covering algorithm based on FP-Tree is much longer because it defining and calculating into the Fuzzy set in finding of association rules. As a result of the accuracy in the search for association rules on this research, the FP-Growth algorithm has a higher accuracy than the Fuzzy c-Covering algorithm based on FP-Tree because of the confidence value in the FP-Growth algorithm reaches 100%. The highest association rules that the FP-Growth algorithm obtained are risoles rahmat, tahu isi emly, pastel bihun susi with support value 0.023%, and confidence value 100%. Meanwhile, the Association rules obtained by Fuzzy c-Covering algorithm based on FP-Tree are aqua 600ml, nasi telur balado siska, tahu bakso siska with a support value of 0.05% and 21% confidence value.

5.2 Suggestions

Suggestions for further research development:

1. Research can be done using transaction data for more than three years to get more accurate results.
2. The application of data mining with the Time Series Analysis method to predict the amount of inventory of items and profits to be gained in the future.

6. REFERENCES

- [1] Adinugroho, S. and Sari, Y. A. (2018). Implementasi Data Mining menggunakan Weka. Malang: UB Press.
- [2] Vlandari, R. T. (2017). Data Mining Teori dan Aplikasi Rapidminer. Yogyakarta: Gava Media.
- [3] Sepri, D. and Afdal, M. (2017). Analisa dan Perbandingan Metode Algoritma Apriori dan FP-Growth untuk Mencari Pola Daerah Strategis Pengenalan Kampus Studi Kasus di Stkip Adzkia Padang. Sistem

Informasi Kaputama, 1(1). 2548-9712.

- [4] Astrina, I., Arifin, M. Z., and Pujianto, U. (2019). Penerapan Algoritma FP-Growth dalam Penentuan Pola Pembelian Konsumen Pada Kain Tenun Medali Mas. *Jurnal Matrix*, 9(1).
- [5] Mayilvaganan, M. and Kalpanadevi, D. (2018). Comparison of Apriori, FP-Tree Growth and Fuzzy FP-Tree Growth Algorithm for Generating Association Rule Mining of Cognitive Skill. *Engineering Research and General Science*, 6(2). 2091-2730.
- [6] Purba, P. C. and Siswanto, E. (2015). Implementasi Algoritme FP-Growth untuk Market Basket Analysis dalam Menentukan Product Bundling.
- [7] Budhi, G. S., Lim, R., and Prayitno, O. (2005). Penggunaan Metode Fuzzy c-Covering untuk Analisa Market Basket Pada Supermarket. *Jurnal Informatika*, 6(1).
- [8] Angraini, K. N., Indwiarti, and Nhita, F. (2018). Implementasi Algoritma Fuzzy c-Covering untuk Mengetahui Pola Pembelian Pada Data Transaksi Swalayan. *e-Proceeding of Engineering*, 5(3), 8198. 2355-9365.
- [9] Sinthuja, M., Puviarasan, N., and Aruna, P. (2017). Evaluating the Performance of Association Rule Mining Algorithms. *World Applied Sciences Journal* 35 (1): 43-53. 1818-4952.
- [10] Bala, A., Shuaibu, M. Z., Lawal, Z. K., and Zakari, R. Y. (2016). Performance Analysis of Apriori and FP-Growth Algorithms (Association Rule Mining). *International Journal Computer Technology & Applications*, 7(2). 279-293. 2229-6093.
- [11] Rindengan, A. J. (2012). Perbandingan Asosiation Rule Berbentuk Biner dan Fuzzy c-Partition pada Analisis Market Basket dalam Data Mining. *Jurnal Ilmiah Sains*, 12(2).