

Data Mining with Associated Methods to Predict Consumer Purchasing Patterns

Hena Lisnawati

Widyatama University, Bandung, 40124, Indonesia
Email: hena.lisnawati@widyatama.ac.id

Ardiles Sinaga

Widyatama University, Bandung, 40124, Indonesia
Email: sinaga.diles@gmail.com

Received: 11 April 2020; Accepted: 25 June 2020; Published: 08 October 2020

Abstract: Technology is developed and utilized as an honest computer in order that it can provide useful information. With the aim of developing and meeting business objectives, the utilization of sales transaction data in minimarket GP is processed into information or knowledge as a recommendation to ascertain the possible value of purchased simultaneously. This processing uses data mining.

Database buildup in computerized systems is justified by getting added value from this data set. Data mining can predicts trends and therefore the nature of business behavior which is extremely useful to support important deciding.

The algorithm wont to form the association rules during this study is CT-Pro. CT-Pro algorithm may be a development of FP-Growth. The difference is within the second step where FP-Growth creates the FP-Tree arrangement while CT-Pro makes the Compressed FP-Tree (CFP-Tree) arrangement. The CT-Pro algorithm process by analyzing employing a tree system where the things most frequently purchased become root and other items will follow the basis. The CFP-Tree process will provide levels for every transaction and facilitate mining results.

CT-Pro algorithm implementation with CFP-Tree arrangement applied to data mining systems is in a position to research sales data for 3 months, namely January 2020 – March 2020 with a complete data record of 1.303 and 320 sales transactions at minimarket GP become information or knowledge. The results of this study are the relationship between the tendency of products that are bought together based on categories in a kind of percentage to be used as a recommendation in structuring the position of items that are mutually frequent in certain categories.

Index Terms: Data Mining, Algorithm CT-Pro, CFP-Tree, Sales Analysis, Mining Results

1. Introduction

Technology as a means to meet human needs. And one of its uses is the application of data mining as a business need. Data mining can be used to dig up data so as to obtain information that can support and improve business processes. The data used in this study to explore data is a database of sales transactions at minimarket GP. A database is an organized collection of logically related data[1].

A. Background

In the 90s there are “piles” of data within the fields of science, business and government. The facility of knowledge technology to collect and storing various kinds of data doesn't leave the facility to research, summarize and extract “knowledge” from the data.

The rapid development of data mining (DM) in conjunction with the event of data technology that allows large amounts of data to accumulate.

Data mining explores database to hunt out hidden patterns, trying to seek out predictive information with can be forgotten by business people. While business people have must utilize data warehouses that are already owned. The sales database continues to accumulate each day.

In the minimarket GP sales transaction data remains left unused. While the arrangement of products is additionally still one manually by watching the tendency of the stock of products that run out first. This makes it difficult within the process of structuring the grouping of categories of products that nee an extended time. With the concept of data mining is predicted to be able to use large data so on produce useful information for the company.

Information is explained as something that results from processing data to be easily understood and meaningful that describe an event and thus the facts that exist. The concept of data mining is predicted to be able to use such large data to provide useful information for the company.

Predicting consumer purchasing patterns during by a method of obtaining information which will be used considerably to support business processes, one among which is to work out the arrangement of adjacent goods in accordance with consumer interest within the tendency to get goods A and goods B purchased simultaneously.

In data mining techniques want to see the connection between data using the association method. Association rule mining takes part in pattern discovery techniques in knowledge discovery and data mining (KDD)[2]. The matter of mining association rules over market basket analysis was introduced in[3] i.e. finding associations between the things that are present within the transaction from the database. Association rule is essentially depends upon the generation of the frequent items within the first phase[4-6]. The database could also be from any retail shop, medical, or from the other applications[7].The association rule consists of several algorithms, including the apriori algorithm, the frequent Pattern-Growth (FP-Growth) algorithm and the Compact Tree-Apriori (CT-Pro) algorithm. An algorithm is presented to get all class association rule and to create an accurate classifier[8].

A large number of algorithms were proposed by many researchers for generation of frequent item sets, firstly apriori and apriori like algorithms are proposed[4,5,9].

According to the new arrangement called the compressed FP-Tree (CFP-Tree) and also the implementation of a replacement FIM algorithm called CT-PRO which was first introduced in[10]. CFP-Tree cohesiveness with FP-Tree is reported at several levels of support in various datasets generated using synthetic data generators[11]. CT-PRO performance was compared with a apriori[4,12], FP-Growth[3], and OP[3].

Based on research on frequent implementations of item mining sets[13], the two best algorithms (LCM and KDCI) from Fimi Repository 2003 and compared their performance with CT-Pro[10]. This shows that CT-Pro is performed better than others[10].

Based on the outline above the prevailing solutions to mountains of knowledge that are often overlooked are the utilization of knowledge mining by applying the association method. The algorithm used in that the CT-Pro algorithm because the choice of the CT-Pro algorithm within the association method is taken into account the simplest so far. The info utilized in this research is sales transaction data. From the resulting pattern will get information or knowledge.

Information or knowledge obtained within the sort of the worth of the tendency of products purchased simultaneously within the sort of a percentage supported 3 (three) main parameters, namely the minimum value of support, minimum confidence and therefore the number of things from the required sales transaction data.

B. Formulation of the Problem

Based on the background of the research, a problem was found, namely how to apply the data mining method with the CT-Pro algorithm to predict consumer purchasing patterns. And how to interpret the purchase patterns produced into information or knowledge.

C. Scope of Problem

In order not to be biased from the target to be targeted or to be more targeted, research is restricted to the scope of the association rules with the info used is that the data on sales transactions. And research doesn't discuss decision support systems or sales information systems.

D. Research Purposes

The purpose of this research is to use data mining method with the CT-Pro algorithm to predict consumer purchasing patterns in minimarket GP and interpret patterns that are produced to be information or knowledge that's beneficial to companies so on increase profits.

E. Research Methodology

In the research methodology plays a crucial role in getting objective and valid data. This is often intended to unravel the issues that are formulated. During this study researchers used quantitative methodology with the sort of experimental research.

Quantitative methods can perform several tasks consistent with the stress of researchers, namely watching comparisons, knowing relationships, seeing trends, doing grouping and simplifying variables.

To try to do this task two main things are needed, first may be a measurement tool and second is an analysis tool. Analysis tools are often likened to a knife which will shape/ carve an object to be ready to see the results (conclusions).

Associated with the background of quantitative research methods are selected and thought of the simplest because the matter may be a clear start line namely the existence of sales transaction data to seek out patterns consumer purchases.

For this reason, during this study arrange the subsequent steps:

- Data source

The data utilized in the study are data obtained from data collection and direct interviews associated with the utilization of purchase transaction data to work out sales stands.

- Data collection techniques

Data collection techniques are administered during a number of the way including: collecting data, i.e. requesting sales transaction data on to the minimarket GP manager who wants to see the info. Then the interview, may be a dialogue conducted by the interviewer to get information from the interviewer. And literature study, is to review the concepts of the Association Rule method found in several sources of literature. Literature sources are often within the sort of textbooks, papers, websites, blogs, journals, and former system technical documents.

- Data validation

Data that has been collected and recorded in research activities must be tried to be true.

- Data analysis

After the info is out there from the results of knowledge collection, processing is then performed. The steps of the analysis work consist of; the pre-processing stage is that the stage of remodeling data into a format that's suitable for analysis. Then the info that has been available in mining a consistent with the chosen algorithm is that the CT-Pro algorithm. Furthermore the mining results are implemented in order that the share results are obtained.

2. Supporting Theory

A. Data Mining

Data mining is that the activity of finding interesting patterns of data in large amounts of data which will be stored in databases, data warehouses, or other information storage [14]. Data mining is related to other fields of science, like database systems, data warehousing, statistics, machine learning, information retrieval, and high-level computing[14].

Data mining may be a process that uses statistical techniques, mathematical techniques, synthesis of intelligence, machine learning in extracting and identifying useful information and knowledge associated with a spread of (large) database[15].

Data mining is an iterative process within which progress is defined by discovery, through either automatic or manual methods[16].

Data mining are often interpreted as a process of managing data by utilizing large amounts of data to get information that's more understandable in order that it can help business processes.

There are several other terms that have an equivalent meaning as data mining, namely knowledge discovery in databases (KDD), knowledge extraction, data/ pattern analysis, business intelligence and data archeology and data dredging[17].

In addition, data mining is supported by other sciences like neural networks, pattern recognition, spatial data analysis, image databases and signal processing[14].

The purpose of data mining is to get relationships or patterns which may provide useful indications[14].

Data mining is split into several groups supported the task which will be done, namely: description, wont to describe patterns and trends contained within the data. Classification, an example of a case is to diagnose a patient's illness to urge into which category of disease. Estimates, almost an equivalent because the classification, except the target variable estimates are more numerical than towards categories. Prediction, almost an equivalent because the classification and estimation except that within the prediction the worth of the results are going to be available within the future. Clustering, may be a grouping of records, observations or attention and form a category of objects that have similarities. Association which determines the attributes that appear at just one occasion. within the business world it's more commonly called basket analysis.

B. CT-Pro Algorithm

CT-Pro algorithm is one among the event algorithms of FP-Growth. It's based upon the compact tree structure[18-20]. The differences are within the second step where FP-Growth makes FP-Tree. The utilization of prefix tree itself was introduced first in[21]. The performance gain from using variants of the prefix tree for representing transactions was previously demonstrated in[11,22,23]. While CT-Pro makes Compressed FP-Tree (CFP-Tree) i.e. compact frequent pattern tree in order that all the things of the transactions are often be represented within the main memory[10,24,25]. At the mining stage the CT-Pro algorithm also uses a bottom-up approach where items within the header table and CFP-Tree are scanned from the littlest to the most important number[10].

CT-Pro algorithm has three stages namely[10]:

- Identify the frequency of things, each data is chosen supported a predetermined minimum support, in order that frequent table items are obtained. Then each item is counted as a frequency of occurrence decreasingly leading in a global item table.
- Building a CFP-Tree, after finding frequent items, a CFP-Tree development is going to be administered. The available frequencies are sorted consistent with global items from the littlest largest value.
- Doing mining, Global item table is sorted from the littlest largest item, because the CT-Pro algorithm works by doing bottom-up mining. Then each global item table is looked for nodes relate to these items, then selected supported the minimum support, the Local Item Table is obtained. From the local item table, CFP-Tree is formed supported the minimum support. Then the items that meet the minimum support are administered mining with the confidence formula.

C. Compressed FP-Tree Data Structure

In this section, a replacement tree-based arrangement, named Compressed FP-Tree (CFP-Tree), is introduced [19]. The following is the presented in Fig. 1.

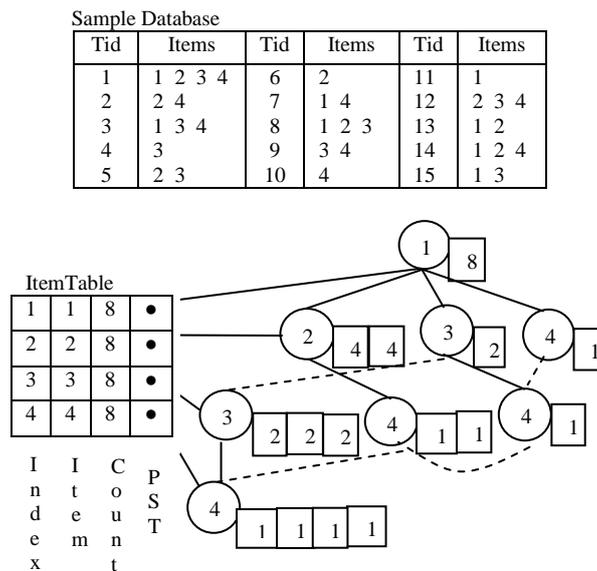


Fig. 1. Compressed FP-Tree

CFP-Tree is defined as follows[10]:

- CFP-Tree consists of a tree that has a root that represents the index of the item with the highest level of occurrence and a collection of subtree as a child of root.
- If $I = \{i_1, i_2, \dots, i_k\}$ may be a collection of frequent items during a transaction, the things within the transaction are going to be entered into the CFP-Tree ranging from the basis subtree which that the i_1 within the header table.
- The root of the CFP-Tree is that the level 0 of the tree.
- Each node in CFP-Tree has 4 main fields namely item-id, parent-id, count which is that the number of things therein node, and therefore the level that shows the structure of the data tree at that node starts from the things contained within the header table with an equivalent level. Found within the CFP-Tree.

D. Lift Ratio

Lift ratio may be a measure to go age the strength of an association rule obtained through a comparison between the confidence of a rule and therefore the value of benchmark confidence. Benchmark confidence itself may be a comparison between the mount of all items that become a consequent (part then) of the entire number of transactions. To calculate the benchmark confidence value, a formula which will be seen within the equation is used:

$$Benchmark = \frac{NC}{N} \tag{1}$$

Information:

NC= Number of transaction with consequent items

N= Number of database transactions

Meanwhile, to calculate the value of the lift ratio uses the formula within the following equation:

$$\text{LiftRatio} = \frac{\text{Confidence (A,C)}}{\text{Benchmark Confidence (A,C)}} \quad (2)$$

An association rule is claimed to point out a benefit if the lift ratio value of the rule is bigger than 1 and therefore the higher the lift ratio value, the greater the association strength.

3. Data Analysis

A. Pre-Processing

At this stage, it's done by compiling an inventory of entities and their required attributes supported system requirements.

Following are the initial steps before data mining:

- Data Cleaning

At this stage, the info obtained comes from sales transaction data in Microsoft Excel format.

The cleaning process is administered by checking the transaction data that has noise and data that are inconsistent or irrelevant then eliminates them.

The following sales transaction data is presented in Table 1.

Table 1. Sales Transaction Data

TID	Date Transaction	Branch	C. Cashier	Category	Code Items	Items	Total
0000I00001	01/01/2016	TBB	001K	Food	MC1012	Chips Ahoy Chocolate Chips Cookies Choco	1
0000I00001	01/01/2016	TBB	001K	Food	MS1006	Chitato Snack Potato Chips Indomie Kornet 55g	1
0000I00001	01/01/2016	TBB	001K	Food	MP1001	Happydent White Chewing Gum 30'S Mint 84g	1
0000I00001	01/01/2016	TBB	001K	Food	MM1033	Indomie Goreng Pedas 80g	1
0000I00002	01/01/2016	TBB	001K	Food	MM1034	Indomie Mie Goreng Ayam Bawang 85g	1
0000I00002	01/01/2016	TBB	001K	Drinks	MR2044	Pocari Sweat Soft Drink 330ML	1
0000I00003	01/01/2016	TBB	001K	Drinks	MR2045	Pororo Minuman Berperisa Blueberry 235ML	1
0000I00003	01/01/2016	TBB	001K	Food	MS1003	Pringles Potato Crisps Original 110 g	1
0000I00003	01/01/2016	TBB	001K	Food	MM1011	Simba Cereal Choco Chips Coklat 55 g	1

In the table there's 2 records, which has no value within the TI field in order that the record is omitted.

- Selection Data

The data within the database aren't needed within the table are going to be deleted, therefore only suitable data to be analyzed is taken from the database. The subsequent attributes which will be retrieved and used for data mining are TID, Category and Items. The following results of data selection in Table 2.

Table 2. Selection Data

TID	Category	Items
0000I00001	Food	Chips Ahoy Chocolate Chips Cookies Choco
0000I00001	Food	Chitato Snack Potato Chips Indomie Kornet 55g
0000I00001	Food	Happydent White Chewing Gum 30'S Mint 84g
0000I00001	Food	Indomie Goreng Pedas 80g
0000I00002	Food	Indomie Mie Goreng Ayam Bawang 85g
0000I00002	Drinks	Pocari Sweat Soft Drink 330ML
0000I00003	Drinks	Pororo Minuman Berperisa Blueberry 235ML

- Transformation

Data preparation must be done to regulate the tools used with the available data formats. Therefore the transformation of knowledge within the design of this technique is to transfer data from Microsoft Excel file format to SQL (Structured Query Language), according to the database tools used in this study, MySQL.

B. Design Process

The system design process uses a functional approach that's represented employing a flowchart. This flowchart shows the physical the processes and data flow within the data mining system which will be created.

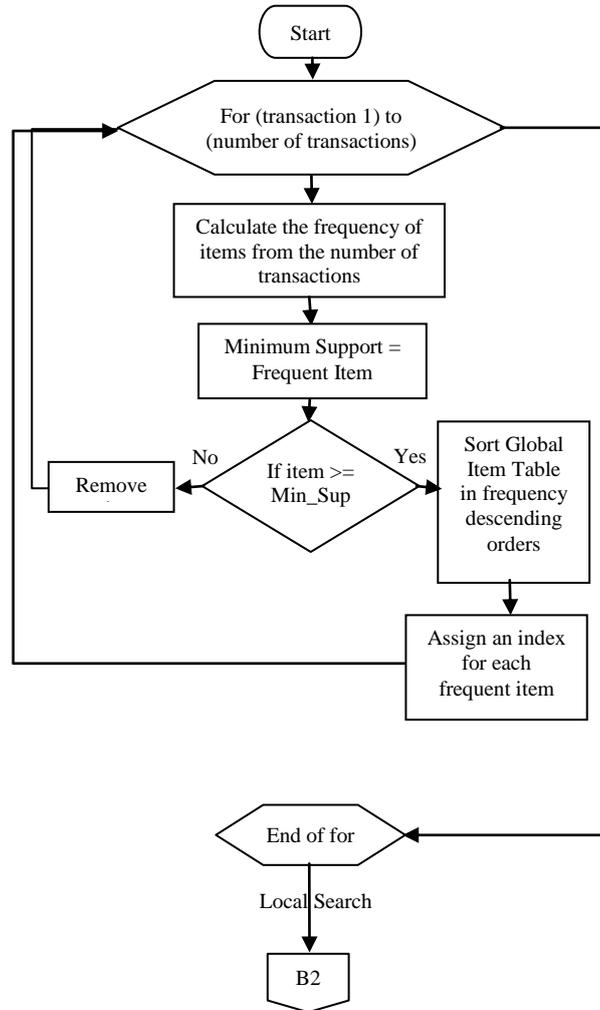


Fig. 2 Flowchart Algorithm CT-Pro

In this section the system flowchart in Fig. 3. indicates that the mining process starts with the step on finding or preparing data then after the info is out there, the edge for input support is minimum support and minimum confidence. If you're getting to display mining results by category select the item category you would like then display the info if it doesn't display the info directly. The subsequent figure may be a general system flowchart in Fig. 3:

The algorithm used to form association rules is the CT-Pro algorithm. The following algorithm flowchart in Fig. 2:

For example, the mining process flow uses 10 transaction data with a minimum support count of 20% and a minimum confidence of 50% presented in Table 3.

The CT-Pro algorithm calculation process stars with counting the number of times the item appears from all transaction. The results are shown in Table 5.

Each item is selected based on a minimum support set of 20%. If the number of items is below the minimum support then the item is removed and can then return to the search process. But if the items above the minimum support exist then the process of forming of global header table is carried out. Formation of global header table tables by means of frequent data sorted from which has the largest to the smallest frequency. And set an index for each frequency of the item. The results are shown in Table 4.

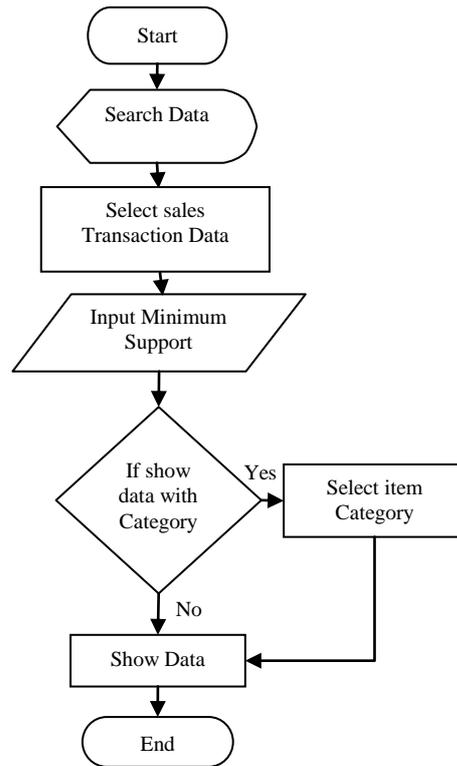


Fig. 3 Flowchart Sistem

Table 3. Data Sample

TID	Items
1	1, 2, 3, 4
2	5, 6
3	1, 7, 4
4	1, 2, 4
5	1, 2
6	2, 8, 9
7	3, 7
8	1, 7, 5
9	2, 3
10	1, 4, 6

Table 4. Global Header Table

Index	Items	Frequents Items
1	1	6
2	2	5
3	4	4
4	3	3
5	7	3
6	5	2
7	6	2

Table 5. Frequent Items

Items	Frequency
1	6
2	5
3	3
4	4
5	2
6	2
7	3
8	1
9	1

Table 6. Local Frequent Items

Index	TID	Items (in index)
7	2	6,7
	10	1,3,7
6	2	6,7
	8	1,5,6
5	3	1,3,5
	7	4,5
	8	1,5,6
4	1	1,2,3,4
	7	4,5
	9	2,4
3	1	1,2,3,4
	3	1,3,5
	4	1,2,3
	10	1,3,7
2	1	1,2,3,4
	4	1,2,3
	5	1,2
	6	2
	9	2,4

Next follows the CT-Pro flowchart (advanced) of the CT-Pro algorithm in Fig. 4:

The CT-Pro flowchart algorithm (advanced) shows that the CT-Pro algorithm works by bottom-up mining so that local frequent item searches are started from the last index. Then select the transaction that contains the index.

The following are the local search results seen in Table 6.

Table 7. Local CFP-Tree

Index	Item (in Index)	Frequency
0,1	1,5	2
0,1	2,4	2
0,1	1,3	4
0,1	2,3	2
0,1	1,2	3

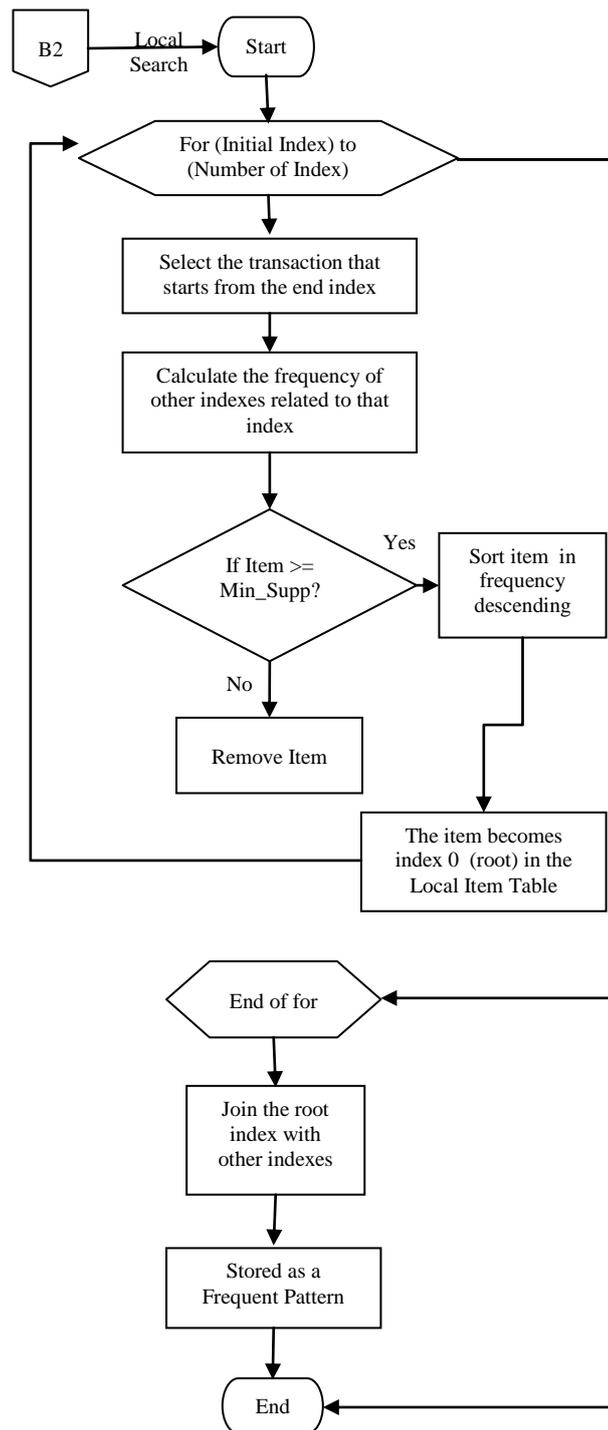


Fig. 4. Flowchart Algorithm CT-Pro (advanced)

Furthermore, if the frequency value of the items obtained is below the minimum support, then remove these items. But if the value can be supported, then the items are arranged from the smallest largest frequency. And with the largest frequency item being index 0 (root). If there are items other than index 0, then combine them to get frequency patterns. The following results are presented in Table 7.

From local CFP-Tree, a frequent pattern is formed according to the mining item. The results are shown in Table 8.

Table 8. Frequent Pattern

No.	Items	Frequency
1	1, 7	2
2	2, 3	2
3	1, 4	4
4	2, 4	2
5	1,2	3

Furthermore, the final stage of the CT-Pro algorithm is to test the patterns which are found whether it can really be used to provide item recommendations. To determine whether a rule is strong or not, the confidence formula is used.

$$Conf (x \Rightarrow y) = \frac{\sum x, y}{\sum x} \tag{3}$$

The following Fig. 5. flowchart confidence value calculation:

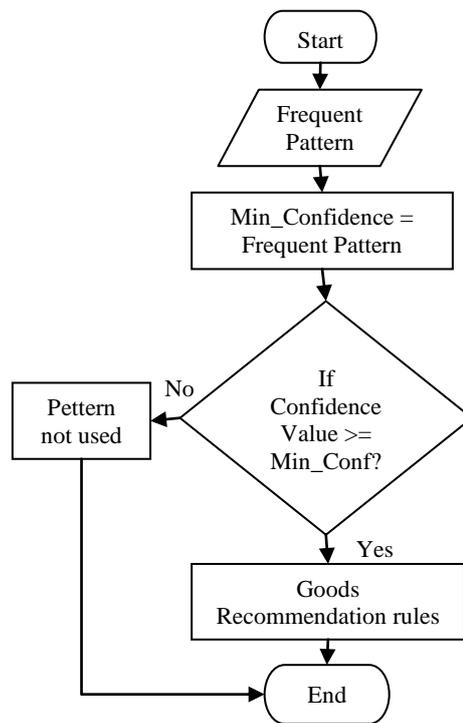


Fig. 5. Flowchart Confidence Calculation

In this section Fig. 5. shows the final process of the CT-Pro algorithm starting with frequent patterns, each item that meets the calculation is calculated. Each item is selected based on a minimum confidence set of 50%. If it meets the minimum confidence, each item concerned is used as knowledge.

In addition to knowing the value obtained is feasible or not, the lift ratio calculation is performed to show the benefits of the associated rules obtained. If the ratio value is greater than 1 and the greater the ratio value, the greater the association strength.

Following are the results of calculating the confidence values for some of the patterns presented in Table 9.

Base on the results of calculations, patterns that meet the minimum confidence value become the rule and used as the basis for providing recommendations.

Table 9. Mining Result

No.	Items	Min. Confidence (...%)	Lift
1	1, 4	67	1.7
2	1, 2	50	1

B. Implementation Mining Results

In this study, before testing the data, what is done is to prepare the purchase transaction data. The data provided is 320 purchase transaction data with 5 categories of goods.

Implementation of data mining in analyzing purchase transactions using association rules with the CT-Pro algorithm. CT-Pro algorithm is the main process of association rules starting with identifying the frequency of items, forming a CFP-Tree and then the mining process.

The following settings for mining that have been determined:

- Minimum support = 0,625%

The frequency of occurrence is worth $2/320 * 100$ or 2 transactions out of 320 transactions, which means that goods purchased less than 2 times will not be processed.

- Minimum confidence = 13,333%

The confidence value of the rule generated is based on a calculation of 2.

The following data presented in Table 10. Is association data mining results:

After the method of calculating the support and confidence of all combinations successfully obtained, confidence rule association value means the products within the rules are always purchased in pairs and from the worth of the support which will be seen there are around 347 record out of 320 transactions.

Tho show the advantages of the related rules obtained are often seen from the ratio value. If the wort of the ratio is bigger than 0.1 and therefore the greater the wort of the ratio the greater the strength of the association.

With the aim of the analysis results as a recommendation regulating the position of the arrangement of the products in accordance with the categories of products desired, info from the results of a comprehensive analysis are often grouped of products purchased together supported the category.

The following may be a comparison to display data by different categories in Table 11.

The results of the comparison show that the tendency of consumers to shop for goods within the food category is that the most desirable as many as 85 records from complete of 320 transaction.

Thus goods that are purchased simultaneously within the food category are often placed on the brink of the beverage items category and by the position of things that are frequency in one another therein category or also can adjust the position of the drug category between food and drinks in order that the chance for consumers to shop for other items within the drug category is its position is found between food and drinks become bigger.

Table 10. Association Rule Results from Mining

Number	Association Rules	Support	Confidence	Ratio
1	Mony Cin Cau Drink Honey 300ML KlqTop Coffee White Coffee White Coffee 10>	0,652	100	1,534
2	Pokka Drink Green Tea Jasmine 450ML, Mony Cin Cau Drink Honey 300ML Klq, '	0,652	100	1,534
3	Sebamed Clear Face Cleansing Foam, Shinzhu'l Body Lotion Skin Lightening Ayu	0,652	100	1,534
4	Mony Cin Cau Drink Honey 300ML Klq, Pokka Minuman The Lemon 350ML	0,652	100	1,534
5	Mony Cin Cau Drink Honey 300ML Klq, Pokka Minuman The Lemon 350ML, Top	0,652	100	1,534
6	Mustika Ratu Peeling Mundisari 125 G, Mustika Ratu Pembersih 2 In 1 Whitening I	0,652	100	1,534
7	Ultra Susu Cair Uht Steril Chocolate 1000ML, Kiranti Sehat Datang Bulan 150ML	0,652	66,667	1,023
8	Good Day Kopi Istant 3 In 1 Chococinno 10X20G, Ultra Susu Cair Uht Steril Chocc	0,652	66,667	1,023
9	Good Day Kopi Istant 3 In 1 Chococinno 10X20G, Ultra Susu Cair Uht Steril Chocolate 1	0,652	66,667	0,711
10	Kacang Koro Orignal 140g, Monde Biscuit Piebis 190G	0,652	66,667	0,711
11	Kao Attack Detergent Powder Plus Softener, Kacang Koro Orignal 140g, Monde E	0,652	40	0,427
12	Kao Attack Detergent Powder Jaz 1 Semerbak, Kao Attack Detergent Powder Plt	0,652	33,333	0,356
13	Jam Pineapple 225g, Kacang Koro Orignal 140g	0,652	66,667	0,711

Table 11. Percentage comparison of the number of paired items by category

Category	Frequency	Percentage
Food	85	26,562 %
Drinks	82	25,625 %
Cosmetics	29	9,0625 %
Drugs	6	1,875 %
Soap	4	1,25 %

4. Conclusion

From the results of the research conducted, the conclusions that can be drawn based on the discussion and implementation in the sale of goods in purchase transactions using the CT-Pro algorithm are the number of goods and goods sold at the same time.

This can be determined by determining 3 (three) main parameters, namely the minimum value of support, the minimum value of confidence, and the number of items in the tasted transaction data.

Display data analysis results by category are often used as recommendation to manage the location of things that are often frequent supported close proximity in order that consumers can more easily shop or also can set the position of things that are often frequent with far distances and place frequent other item set among them, in order that the chance for consumers to shop for other items whose position is found between frequent items is bigger.

References

- [1] Jeffrey A. Hoffer, Mary B. Prescott, and Fred R. Mcfadden, *Modern Database Management*, 7th ed.: Prentice, 2004.
- [2] Bharat B. Gupta and Deepak Garg, "FP-Tree Based Algorithms Analysis: FP- Growth, COFI-Tree and CT-PRO," *International Journal on Computer Science and Engineering (IJCSSE)*, vol. 3, pp. 2691-2699, July 2011.
- [3] Jiawei Han, Jian Pei, and Yiwen Yin, "Mining Frequent Patterns Without Candidate Generation," *ACM sigmod record*, pp. 1-12, 2000.
- [4] Rakesh Agrawal and Ramakrishnan Srikant, "Fast Algorithms for Mining Association rules," *Proc. 20th int. conf. very large data bases, VLDB*, pp. 487-499, 1994.
- [5] Jiawei Han and Micheline Kamber, *Data Mining-Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers, 2009.
- [6] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami, "Mining association rules between sets of items in large databases," *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pp. 207-216, June 1993.
- [7] M-L Antonie and Osmar R Zaiane, "Text document categorization by term association," in *IEEE International Conference on Data Mining*, pp. 19-26, December 2002.
- [8] Bing Liu, Wynne Hsu, and Yiming Ma, "Integrating classification and association rule mining," *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, 1998.
- [9] Qihua Lan, Defu Zhang, and Bo Wu, "A New Algorithm For Frequent Itemsets Mining," *Proceeding on IEEE National Conference on Global Congress on Intelligent Systems*, pp. 360-364, 2009.
- [10] Yudho Giri Sucahyo and Raj P Gopalan, "CT-PRO: A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure," *FIMI*, pp. 212-223, 2004.
- [11] Liu J, Pan Y, Wang K, and Han J, *Mining Frequent Item Sets by Opportunistic Projection*. Canada: Proceedings of ACM SIGKDD, 2002.
- [12] Christian Borgelt, "Efficient Implementations of Apriori and eclat," *Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI), Melbourne, Florida*, 2003.
- [13] Fimi. (2003) Fimi Repository. [Online]. <http://fimi.cs.helsinki.fi>
- [14] Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*.: Elsevier, 2006.
- [15] Efraim Turban, Jay E Aronson, and Ting Peng Liang, *Decision Support Systems and Intelligent Systems*, 7th ed.: Prentice Hall, 2004.
- [16] Mehmed Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed.: John Wiley & Sons, 2011.
- [17] Daniel T Larose, *Discovering Knowledge In Data : An Introduction to Data Mining*. Canada: John Wiley & Sons, Inc., 2005.
- [18] Raj P Gopalan and Yudho Giri Sucahyo, "High performance frequent patterns extraction using compressed FP-tree," *Proceedings of SIAM International Workshop on High Performance and Distributed Mining (HPDM), Orlando, USA*, 2004.
- [19] Yudho Giri Sucahyo and Raj P Gopalan, "CT-ITL: Efficient Frequent Item Set Mining Using a Compressed Prefix Tree With Pattern Growth," *Proceedings of the 14th Australasian database conference-Volume 17*, pp. 95-104, 2003.
- [20] SP Latha and N Ramaraj, "Algorithm for Efficient Data Mining," *Proceeding on IEEE International Computational Intelligence and Multimedia Applications*, vol. 2, pp. 66-70, 2007.
- [21] Ramesh C. Agarwal, Charu C. Aggarwal, and V.V.V. Prasad, "A Tree Projection Algorithm for Generation of Frequent Item Sets," *Journal of Parallel and Distributed Computing (Special Issue on High Performance Data Mining)*, 2000.
- [22] Jiawei Han, Jian Pei, Yiwen Yin, and Runying Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," *Data Mining and Knowledge Discovery: An International Journal, Kluwer Academic Publishers*, vol. 8, pp. 53-87, 2004.

- [23] W. Cheung and O. R. Zaiane, "Incremental mining of frequent patterns without candidate generation or support constraint," *Proceedings of Seventh International Database Engineering and Applications Symposium*, July 2003.
- [24] Jochen Hipp, Ulrich Gützer, and Gholamreza Nakhaeizad, "Algorithms for association rule mining—a general survey and comparison," *ACM sigkdd explorations newsletter*, vol. 2, no. 1, pp. 58-64, June 2000.
- [25] Aiman Moyaid Said, PDD Dominic, and Azween B Abdullah, "A comparative study of fp-growth variations," *International journal of computer science and network security*, vol. 9, no. 5, pp. 266-272, may 2009.

Author's Profile



Hena Lisnawati, obtained a Diploma Degree (A.Md) in Information Management from Bandar Lampung (DCC Bandar Lampung). The writer is currently busy managing his own and family business in the field of trade and personal services. His business in the field of trade was named Gilang Padasuka (retail and wholesale). Whereas in the field of services per person named Joint Partners is a printing and private tutoring.

How to cite this paper: Hena Lisnawati, Ardiles Sinaga, " Data Mining with Associated Methods to Predict Consumer Purchasing Patterns", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.12, No.5, pp. 16-28, 2020.DOI: 10.5815/ijmeecs.2020.05.02