

# MARKETING STRATEGY FOR THE DETERMINATION OF STAPLE CONSUMER PRODUCTS USING FP-GROWTH AND APRIORI ALGORITHM

<sup>1</sup>Winda Widya Ariestya, <sup>2</sup>Wahyu Supriyatin, <sup>3</sup>Ida Astuti

<sup>1,2,3</sup>Faculty of Computer Science and Information Technology Gunadarma University  
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

<sup>1</sup>winda\_widya@staff.gunadarma.ac.id, <sup>2</sup>ayu\_ws@staff.gunadarma.ac.id,

<sup>3</sup>astuti@staff.gunadarma.ac.id

## Abstract

The demand for staple products that vary among customers makes it necessary for the store to determine how the marketing strategy should be. Data mining are known as KDD (Knowledge Discovery in Database) is to digging up valuable knowledge from the data. Research purpose is to identify the right marketing strategy to sales the goods. The marketing strategy is took by analyze how much consumers demand for basic needs. The algorithms used in this research are FP (Frequent Pattern)-Growth and A-priori Algorithm. Finding combinations patterns between item set using the Association Rule. FP-Growth algorithm is an algorithm that been used to determining a set of data in a data set that often appears on the frequency of the item set. the KDD stages study are data cleansing, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. the Testing used RapidMiner software with a minimum confidence value of 0.6 and a minimum support of 0.45. FP-Growth algorithm obtained 5 rule conclusions while Apriori Algorithm obtained 3 rule conclusions. The FP-Growth algorithm make a better decision rules than a priori algorithms in determining of marketing strategies, because it produces more decisions on how the goods sold.

**Keywords:** Apriori Algorithm, Association Rule, Data Mining, FP-Growth Algorithm, Marketing Strategy

This paper is a revised and expanded version of a paper entitled Marketing Strategy for The Determination of Staple Consumer Products using FP-Growth and Apriori algorithm presented at UG Economic Faculty International Conference October 31<sup>st</sup>, 2019, Bekasi, West Java, Indonesia.

## INTRODUCTION

Mining data or attempts to dig up useful information and knowledge in very large databases is called data mining or KDD (Tampubolon, 2013). The growing experience of using data mining techniques. Data Mining is a methodology used in the effective and efficient discovery of existing knowledge in a large-scale data set (Ependi, 2019). Excavating the information algorithm is achieved in the hope that it can aid both majerially and in business processes in the decision-making process (Ependi, 2019). Data mining is carried out in different areas, including business or commerce, education and telecommunications (Tampubolon, 2013). Recommendations for product promotion are

an application model of observation results on customers' circumstances and desires to buy a product. Recommendations for brand marketing use consumer feedback on an item to help consumers choose goods (Triyanto, 2014). This study makes use of a comparison between the algorithm FP-Growth and the algorithm Apriori. In implementations, the algorithm is used to determine the trend of consumer purchases of basic needs goods in order to be used to assess the next marketing strategy. We must obtain information from the two algorithms from the resulting pattern and the best marketing approach can be seen.

The purpose of this study is to analyze and define market strategies as

steps taken by stores to meet basic market needs. The marketing strategy is performed to assess which staple products are most in consumer demand. Stores will only sell goods that customers are searching for and need.

Throughout his research Fajrin (2018) used the FP-Growth Algorithm to help companies understand product buying patterns and sales of spare parts. In science, data mining is used to obtain information that is used as a tool for enhancing business activities. In 2014, Triyanto stated that association rule mining can provide decision support among goods that customers jointly purchase. Using the FP-Growth Algorithm for the association method, the study was able to help determine the appropriate recommendations for brand promotion by generating 3 decision rules.

Ependi (2019) projected inventory of goods with a research in 2019 by processing sales data using the Data Mining method with Apriori algorithm based on the consumer purchase process based on the relationship between the purchased products. Tampubolon (2013) explains that applying a priori algorithms to data mining techniques is highly efficient and can speed up the process of combining patterns of itemset combinations resulting from the selling of medical devices at Pharmacy Kelambir 2 Medan. Shalaby (2015), in retail sales, is using the Apriori Algorithm and the FP-Growth Algorithm. Brand building is one of the most effective marketing strategies used to create inventory by establishing credibility or communication at discounted prices between unsolicited and requested goods. The way it was built using the lowest price by implementing a recommendation system and increasing customer trust by following the buying habits. Research on market analysis of baskets using FP-Growth is proposed to determine the availability of

planning of goods. To find out patterns of consumer spending at Berkah Mart, Pekanbaru, the application of the FP-Growth algorithm can produce goods and informative association rules (Ilham, 2018). Budihartanti's (2013) research has developed the three highest rules that can be used to encourage goods for sale. According to the level of demand, businesses must be responsive to the level of sales of each product.

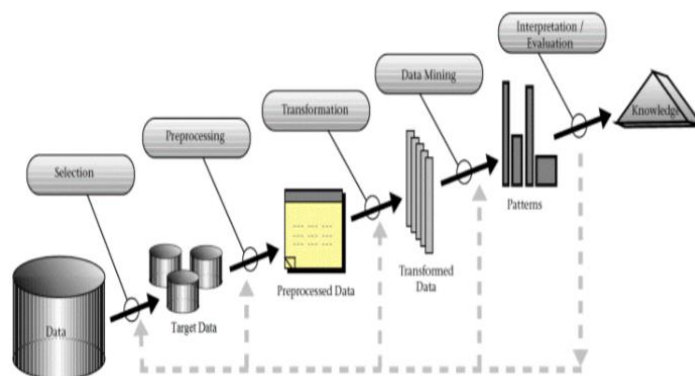
## LITERATURE REVIEW

### **KDD (Knowledge Discovery in Database)**

KDD is defined as the extraction from a data set of potential, implied and unknown information.

The process of knowledge discovery involves the results of the data mining process (the process of extracting data patterns), then the results are accurately translated to easily understood information. KDD itself is defined as a non-trivial process for finding and identifying patterns in data where found patterns are legal, new, useful and understandable (Fajrin, 2018).

The data mining stages are: (Ependi, 2019) (a) Data Cleaning to eliminate noise and data that are inconsistent, (b) Data integration, the combination of data sources process, (c) Data Selection, the process of analyzing relevant data. (d) Data Transformation, the method of data transformation and consolidation into a form appropriate for the development of description or integration of mining. (e) Data Mining, the initial phase through which methods of analysis are used to extract patterns of data. (f) Pattern Analysis, the method of finding interesting patterns that are based on a certain size as a basis of knowledge, (g) Presentation of knowledge, process of visualization techniques and presentation of knowledge used to display knowledge to users.



**Figure 1. Stages of the KDD Process in Data Mining**  
Source: Ependi, 2019

### Data Mining

One of the problems with the approval of Data Mining is that the determinants of Data Mining inherit most characteristics and techniques from already developed fields.

Starting from a number of scientific disciplines, Data Mining encourages the refinement of traditional techniques in order to improve them : (Fajrin, 2018): (a) Very large data volume , (b) High data volume (c) Heterogeneous data and various properties KDD is defined as the extraction from a data set of potential, implied and unknown knowledge. Knowledge Process Discovery in Database involves the results of the data mining process (the process of extracting the trend of a data pattern), then accurately converting the results into easily understood information (Tampubolon, 2013).

### Association Rule

Association Rule or ARM (Association Rule Mining) is a one-subprocess technique or system used to process data. First, a range of strategies is called a step to find frequent items, and second is to determine the association's mining rules. The principle of using items, which can then become a positive association rule or negative association rule, can be extracted during this sub-method. The basic method can be done in two stages in the process of association analysis (Ependi, 2019): (a) High frequency pattern review: Use the formula to find the value of your minimum requirements from the value of the item set support, (b) Establishment of Associative Rules: After all patterns with high frequency are found, calculate the Confidence value, Then to find two or more itemsset:

$$Support (W) = \frac{\sum Transaction\ containing\ W}{\sum Transaction\ that\ have\ occurred} \quad (1)$$

$$Support (W, X) = \frac{\sum Transaction\ containing\ W\ and\ X}{\sum Transaction\ that\ have\ occurred} \quad (2)$$

$$Confidence (W > X) = \frac{\sum Transaction\ containing\ W\ and\ X}{\sum Transaction\ containing\ W} \quad (3)$$

Market basket analysis (Association Rule Mining) is a method of data mining

that focuses on finding patterns of purchase by extracting associations or events from

transactional data from a store. In addition, market basket analysis is a good way to support the retail market by offering objective decision support through the interaction between items purchased together by the mining association.

Discovering this partnership can help traders build selling strategies by understanding products that buyers also purchase together, which is very useful because it can enable product recommendations and product promotions to make marketing strategies more suitable. Determining inaccurate patterns of purchasing goods can lead to non-target product recommendations and product promotion policies (Triyanto, 2014).

#### **Algoritma FP-Growth (Fajrin, 2018)**

FP-Growth (Frequent Pattern Growth) is an alternative algorithm that can be used to evaluate the data set that occurs most often in a data set. The FP-Growth algorithm is an Apriori algorithm development. The FP-Growth algorithm corrects the deficiencies of the Apriori algorithm. FP-Growth algorithm is the development of a priori method which is one of the alternatives for determining the data set that most frequently appears in a data set by generating a data tree structure or called the Frequent Pattern Tree (FP-Tree) (Fajrin, 2018).

The concept of tree development is used by FP-Growth to search for frequent itemsets. That's what makes the algorithm of FP-Growth faster than the algorithm of Apriori. The FP-Growth algorithm's feature is that the data structure used is a tree called the FP-Tree. The FP-Growth algorithm can extract frequent item set directly from FP-Tree by using FP-Tree.

FP-Tree is a system of compact data storage. FP-Tree is constructed by mapping each transaction data into each particular FP-Tree path so transactions that have the same element that occur in each mapped transaction, allowing the path to overwrite each other. The more transaction data with the same item, the more efficient is the process of compression with the FP-Tree data structure.

A data tree structure (FP-Tree) will be used to extract frequent item sets using the FP-Growth algorithm. The process of FP-Growth can be divided into three main stages: (a) The conditional pattern base generation process: Conditional Order Base is a sub database with a path of prefixes and an order of suffixes. A conditional sequence base is developed via previously constructed FP-Trees. (b) Step of conditional generation of FP-Tree: At this stage, the support count of each item on each conditional pattern base will be summed up, then each item with a conditional FP-Tree that has a number of support counts higher than the minimum support count. (c) The frequent search phase for items: If the Conditional FP-Tree is a single path, by combining items for each conditional FP-Tree, a frequent item set is obtained. If it is not a single path, then FP-Growth will be generated recursively.

#### **Algoritma Apriori (Ependi, 2019)**

The purpose of the Apriori algorithm is to find the common item set run by a set of data. All items that have k elements, which is called K-item set, are included in the K-iteration (Budihartanti, 2013).

The method that takes place in the Apriori algorithm is to evaluate the regular item set that will be processed by using an iterative or level-wise search approach to obtain the current association rules.

In the Apriori algorithm, the mechanism is as follows: (a) The process by which a candidate item is generated set, (b) Process of data scanning for calculating support, (c) Determine rules for candidates containing possibilities for rules with a support value > minimum support, (d) Determine the value of the antecedent support merged with Table F by comparing the support rule value with the value generated for the preceding support.

## **RESEARCH METHOD**

Data collection approach is the method used in this analysis, namely:

### **Research in the field**

Research is conducted directly to store basic necessities of goods in order to collect primary data.

The techniques used in the processing of information are: (a) Observation, collection of information by direct observation and documentation of all the data required for study, (b) Interview, data collection by asking and responding directly to the parties involved in the field being examined.

### **Research in Literature**

Research by conducting literature reviews to study research-related books. Use the RapidMiner software to learn the tools used in the algorithm application. The steps of the experimental method used in this study are as follows: (a) Methods of data collection, (b) Algorithms Application (FP-Growth and Apriori Algorithms), (c) Evaluation of outcomes, (d) Documentation of experimentation, (e) Comparison of the results of the algorithm. The flow diagram of this analysis can be seen in Figure 2.

## **RESULT AND DISCUSSION**

The KDD approach is used in this study. The KDD levels in this analysis are as follows:

### **Cleaning of data**

The data being processed is the sales data of the basic goods.

Using RapidMiner software, the process of data cleaning or data cleaning sales data on basic needs items. At this stage, data that is problematic due to double data will be cleaned. The dataset used was 620, but data was cleaned up to 577 data.

### **Integration of data**

This stage is achieved by merging data from different sources needed to form the dataset system. The data set used is the data selling goods of basic needs that are processed using the FP-Growth

Algorithm and Apriori Algorithm to extract information (data mining). Figure 3 shows the formation of the dataset for the FP-Growth Algorithm and Apriori Algorithm.

### **Selection of the data**

The data collection process is done by choosing the attributes or fields in the dataset required to perform data mining. The attributes used in this study are the types of commodities that consumers purchase and the names of consumers who purchase.

The dataset is analyzed using RapidMiner technology in a table format in Excel as shown in Figure 4. In the selection process, the data set used amounts to 577 data.

### **Transformation of the data**

System of data conversion if data originates from various table sources. Using data from one table in one database for this analysis.

### **Mining of data**

Upon acquiring a data set suitable for mining needs using RapidMiner, the data mining process is performed.

The operator association rule model is made using the operator format used from the data in Figure 4.

Figure 5 is the FP-Growth algorithm template operator rule algorithm and Figure 6 is the Apriori algorithm association principle

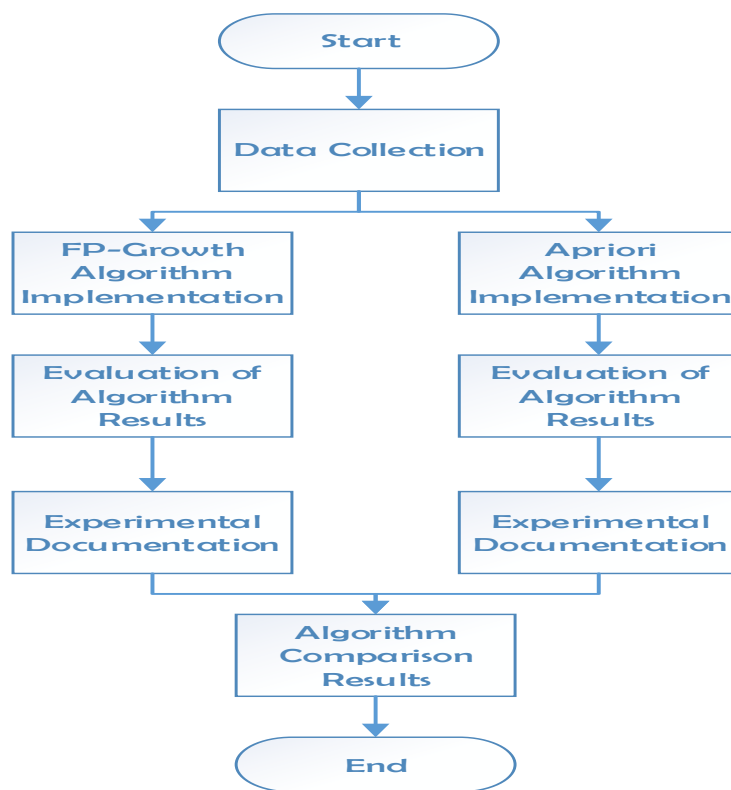


Figure 2. Research phases flow

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	A	B	C	D	E	F	G	H	VARIAN	COUNT		A	MINYAK	
2	1	0	0	1	0	1	1	0	ADFG	2		B	BERAS	
3	0	0	0	1	1	1	1	0	DEFG	3		C	TEPUNG	
4	0	0	1	0	0	1	0	0	CF	3		D	SNACK	
5	1	0	1	0	1	1	0	0	ACEF	1		E	GULA	
6	1	0	0	0	1	0	0	0	AE	1		F	TELUR	
7	1	0	1	0	0	0	0	0	AB	1		G	KOPI	
8	1	0	1	0	1	0	0	0	ABE	1		H	MIE INSTAN	
9	0	0	0	1	1	0	0	1	DEH	8				
10	0	0	1	1	1	0	1	1	CDEGH	3				
11	0	0	1	1	1	1	1	1	CDEGH	7				
12	0	0	0	1	1	1	0	0	DEF	2				
13	0	1	0	1	0	1	1	1	BDFGH	8				
14	0	1	0	0	0	1	1	1	BFGH	4				
15	0	1	0	0	0	1	1	0	BFG	2				
16	1	1	0	0	0	0	1	0	ABG	1				
17	0	1	0	1	0	0	0	0	BD	5				
18	1	1	0	1	0	0	1	0	ABDG	2				
19	0	1	1	0	1	0	1	0	BCEG	6				
20	0	0	1	0	1	0	0	0	CE	3				

Figure 3. Integration of data

	A	B	C	D	E	F	G	H	I	J	K
1	Penjualan	A	B	C	D	E	F	G	H		
2	P1	1	1	1	1	0	0	0	0		
3	P2	1	1	1	1	0	0	0	0		
4	P3	1	1	1	1	0	0	0	0		
5	P4	1	1	1	1	1	0	0	0		
6	P5	1	1	1	1	1	0	0	0		
7	P6	1	1	1	1	1	0	0	0		
8	P7	1	1	1	1	1	1	0	0		
9	P8	1	1	1	1	1	1	0	0		
10	P9	1	1	1	1	1	0	1	0		
11	P10	1	1	1	1	1	0	1	0		
12	P11	1	1	1	1	1	0	1	0		
13	P12	1	1	1	1	1	0	1	1		
14	P13	1	1	1	1	1	0	1	1		
15	P14	1	1	1	1	0	0	0	1		
16	P15	1	1	1	1	0	0	0	1		
17	P16	1	1	1	0	1	1	1	1		
18	P17	1	1	1	0	1	1	1	1		
19	P18	1	1	1	0	1	0	1	0		
20	P19	1	1	1	0	1	0	1	0		
21	P20	1	1	1	0	1	0	1	1		
22	P21	1	1	1	0	1	0	1	1		

Figure 4. Selection of the data

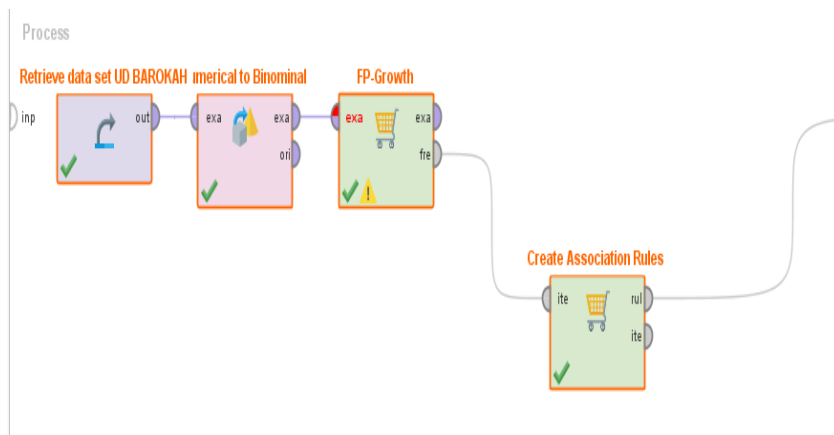


Figure 5. Rule FP-Growth algorithm for model operator association

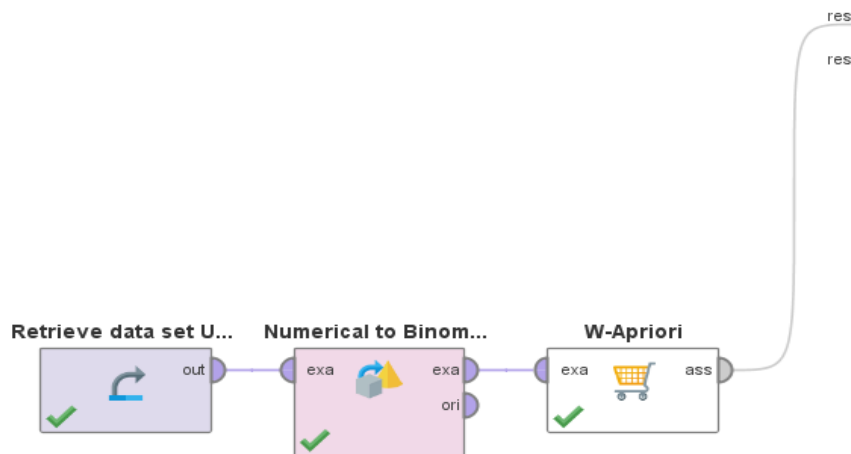


Figure 6. Model operator's association rule Apriori algorithm

The results obtained following the execution of the FP-Growth algorithm for the operator association rule are as follows. From the association regulation, 5 rules for the purchase of goods are obtained: H (Instant

Noodle), G (Coffee), E (Sugar), F (Egg) and C (Flour). Figure 7 is the connectivity of the 5 goods produced, with a value of min trust of 0.6 and min support of 0.45. Figure 9 is the conclusions of products F.

#### Association Rules

```
[E, C, D] --> [F] (confidence: 0.600)
[E, C, B ] --> [F] (confidence: 0.600)
[C] --> [E] (confidence: 0.601)
[H] --> [E] (confidence: 0.603)
[H, E] --> [F] (confidence: 0.604)
[F, D] --> [G] (confidence: 0.604)
[E, D] --> [H] (confidence: 0.607)
[H, G, E] --> [C] (confidence: 0.607)
[H, G, C] --> [F] (confidence: 0.608)
[F, B ] --> [H] (confidence: 0.608)
[E, C] --> [F] (confidence: 0.614)
[G, E, D] --> [C] (confidence: 0.615)
[F] --> [E] (confidence: 0.616)
[G] --> [E] (confidence: 0.617)
[E, F, C] --> [H] (confidence: 0.619)
[G, D] --> [H] (confidence: 0.621)
[H, F, B ] --> [G] (confidence: 0.621)
[C] --> [G] (confidence: 0.621)
[H, G, E] --> [F] (confidence: 0.621)
[E] --> [H] (confidence: 0.622)
[F, C] --> [H] (confidence: 0.622)
[F, D] --> [E] (confidence: 0.625)
[G, C, D] --> [H] (confidence: 0.625)
[E, F, B ] --> [H] (confidence: 0.625)
[G, E] --> [C] (confidence: 0.630)
[G, C] --> [H] (confidence: 0.632)
[H, G] --> [E] (confidence: 0.633)
[E] --> [G] (confidence: 0.634)
[F] --> [H] (confidence: 0.635)
[G, B ] --> [H] (confidence: 0.636)
[B ] --> [E] (confidence: 0.639)
[G, E, C] --> [H] (confidence: 0.639)
[G, E, C] --> [F] (confidence: 0.639)
[E, C, D] --> [G] (confidence: 0.640)
[H, F] --> [E] (confidence: 0.641)
```

The results of the Apriori algorithm association rule operator model are as follows:

```
Minimum support: 0.45 (58 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 3
```



Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 21

Size of set of large itemsets L(3): 9

Best rules found:

1. C=false H=false 99 ==> B =true 63      conf:(0.64)
2. B =true G=false 115 ==> C=false 70      conf:(0.61)
3. B =true E=false 100 ==> C=false 60      conf:(0.6)

### Evaluation of patterns

In the data mining analysis process using the FP-Growth algorithm with a minimum confidence value of 0.6 or 60 per cent and a minimum support value of 0.45 or 45 per cent, 5 patterns of goods transactions with a maximum frequency of up to 5 items were identified. A priori algorithm with a minimum confidence value of 0.6 and a minimum support value of 0.45 found 3

patterns of consumer transactions with a maximum frequency of up to 3 items, all of which are translated into simpler forms so that they can be easily understood. The KDD comparison of the two algorithms used shows that the FP-Growth algorithm was able to provide a better decision rule by obtaining 5 items set for the type of goods that are in high demand. Form of implementation.

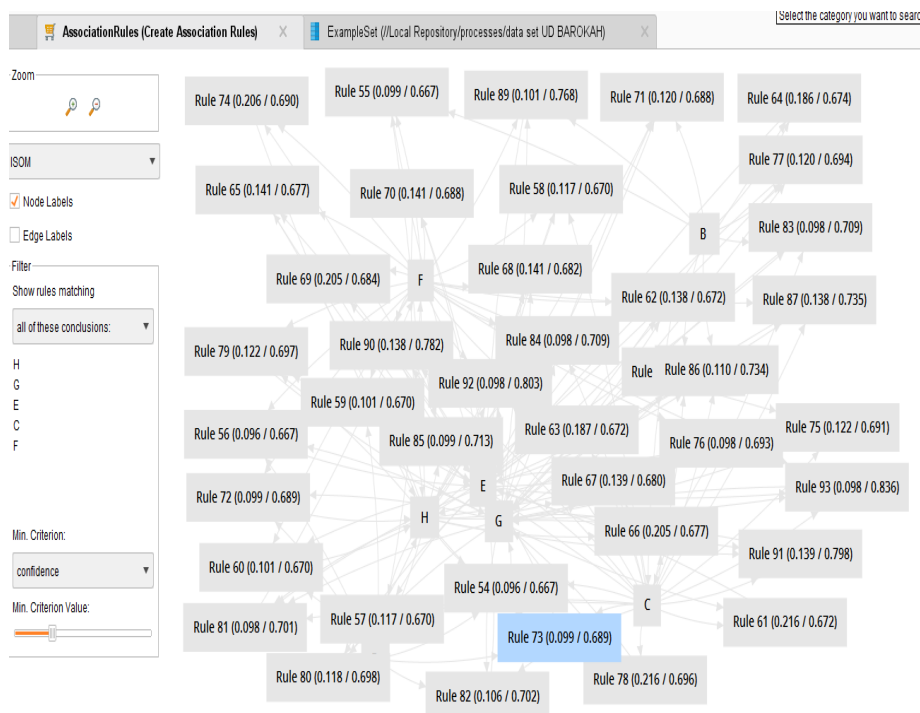


Figure 7. Chart of the relationship between goods on the basis of the value of support and confidence

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
2	E, C	F	0.187	0.603	0.906	-0.434	0.027
10	H, G, E	F	0.141	0.615	0.928	-0.317	0.023
35	G, E, C	F	0.138	0.637	0.936	-0.294	0.026
57	H, E, C	F	0.117	0.670	0.951	-0.232	0.027
81	H, G, E, C	F	0.098	0.701	0.963	-0.181	0.026

**Figure 8. Conclusions Products F**

of the marketing strategy implementation by selling 5 types of staple goods that are much in demand by consumers, namely H (Instant Noodle), G (Coffee), E (Sugar), F (Egg) and C (Flour). These five basic requirements are combined in sales in order to increase the value of product sales. For example, in Figure 9, if consumers buy instant noodles, sugar and flour, consumers will definitely buy eggs like the Rapid Miner rule.

### CONCLUSION AND SUGGESTION

The marketing strategy for buying and selling is something that traders need to do to increase sales of their production quantity. Marketing strategies can be determined on the basis of the different conditions between the location of the market, the number of requests for certain products, the level of consumer demand, the level of consumer income or the number of competitors. This research explores how the marketing strategy that a grocery store will follow in order to determine what staples customers are in great demand. The research was conducted by comparing two algorithms, FP-Growth and Apriori. The algorithm is carried out by following the KDD stage by producing

an association rule as the decision of each algorithm.

The FP-Growth algorithm has a better connection law than the Apriori algorithm. FP-Growth Algorithm generates 5 basic consumer decision-making rules. The peak frequency generated using the FP-Growth algorithm obtains 5 sets of products that can be used by stores to evaluate market strategies. Apriori's algorithm produces only 3 decisions with a maximum frequency of up to three item set, so it is less likely to be used in the cooking strategy. Then the FP-Growth algorithm was able to provide a marketing strategy solution with the number of 5 decision rules produced.

Further development to identify marketing strategies using other KDD algorithms as FP-Growth and Apriori development. Then, with the existing conditions, the algorithm used can produce better decisions and more rules to generate a more variable rule.

### BIBLIOGRAPHY

Budihartanti, C. (2013). Penerapan data mining berdasarkan asosiasi menggunakan algoritma apriori dalam penentuan pola belanja kitchen

- appliances. *Jurnal Techno Nusa Mandiri*, 10(1), 20-28.
- Ependi, U., & Putra, A. (2019). Solusi prediksi persediaan barang dengan menggunakan algoritma apriori (studi kasus: regional part depo auto 2000 Palembang). *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 5(2), 139-145.
- Fajrin, A. A & Maulana, A. (2018). Penerapan data mining untuk analisis pola pembelian konsumen dengan algoritma FP-Growth pada data transaksi penjualan spare part motor. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 5(1), 27-36.
- Mustakim, Herianda, D.M., Ilham, A., Daeng., A. GS, Laumal, F. E., Kurniasih, N., Iskandar, A., Manulangga, G., Iswara, I.B.A.I & Rahim, R. (2018). Market basket analysis using apriori and FP-Growth for analysis consumer expenditure patterns at Berkah Mart in Pekanbaru Riau. In *Journal of Physics: Conference Series*, 1114(012131), 1-9.
- Shalaby, A.H., el letif, D. M., Rezk, A., & Hagag, M. (2015). Suggested marketing strategy using apriori and FP-Growth algorithms in retail sales in Egypt. *International Journal Of Computers & Technology*, 14(11), 6190-6200. DOI: <https://doi.org/10.24297/ijct.v14i11.1807>.
- Tampubolon, K., Saragih, H., Reza, B., Epicentrum, K., & Asosiasi, A. (2013). Implementasi Data Mining Algoritma Apriori pada sistem persediaan alat-alat kesehatan. *Informasi dan Teknologi Ilmiah (INTI)*, 1(1), 93-106.
- Triyanto, W. A. (2014). Association Rule Mining Untuk Penentuan Rekomendasi Promosi Produk. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 5(2), 121-126.