

# COMPARISON OF MARKET BASKET ANALYSIS METHOD USING APRIORI ALGORITHM, FREQUENT PATTERN GROWTH (FP-GROWTH) AND EQUIVALENCE CLASS TRANSFORMATION (ECLAT) (CASE STUDY: SUPERMARKET “X” TRANSACTION DATA FOR 2021)

Rina Wahyuningsih<sup>1</sup>, Agus Suharsono<sup>2\*</sup>, Nur Iriawan<sup>3</sup>

<sup>3</sup>Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

E-mail : [agus\\_s@statistika.its.ac.id](mailto:agus_s@statistika.its.ac.id)

**Abstract:** The retail industry continues to grow and develop in Indonesia. The retail sector as a provider of goods used in everyday life has long started digital transformation in its business. Digital technology helps the retail industry collect valuable customer data. Business analytic is the use of data, information technology and statistical analysis, which is used to obtain information about a business and make decisions based on facts. Business analytic turns data into steps or actions in the context of making business decisions. Consumer needs and purchasing behavior can be predicted with big data-based technology. Association Rule is a technique in data mining to find the relationship between items in an item set combination. One of the utilizations of the association rule method is market basket analysis. Algorithms that can be used to analyze consumer purchasing patterns include the Apriori algorithm, Frequent Pattern Growth (FP-Growth) which represents a database structure in a horizontal format, and the Equivalence Class Transformation (ECLAT) algorithm which represents a vertical data format. In addition, this research will first analyze the complexity of the algorithm based on the time complexity in running the algorithm. This analysis uses these three algorithms, which are applied to Supermarket "X" transaction data in 2021, namely 136,202 transactions. The measure of goodness that is used to find out the best algorithm uses support and confidence values. The results show that the ECLAT algorithm is the most superior algorithm compared to the others based on the execution time required by the algorithm. The support value used in forming associations in the ECLAT algorithm is 1%, resulting in 19 rules. From the results of these rules, the highest support value was generated by the purchase of Indomie goreng special and Indomie ayam bawang, where as many as 1,362 shopping transactions bought these two items together or 2.71% of the total transactions.

**Keyword:** Retail, Big Data, Association Rules, Apriori, FP-Growth, ECLAT

## INTRODUCTION

The retail industry continues to grow and develop in Indonesia. This is indicated by the existence of traditional markets which are starting to be displaced by the emergence of various types of modern markets, such as supermarkets and shopping centers. Shopping centers provide various commodities for community needs, ranging from primary needs, secondary needs to tertiary needs. In contrast to shopping centers, most supermarkets sell more commodities related to consumer goods as the main commodity. The trade industry is a sector that continuously provides a concrete contribution to the

composition of Indonesia's economy. Retailing is a type of business that entails the sale of products or services to individual customers in small quantities or units. Retail consumers purchase products or services with the intention of utilizing or using them personally, and not with the aim to resell them. Indonesia is experiencing a rapid growth in the retail industry. The rapid development of the retail business is also supported by Indonesia's high population, so that Indonesia is the most potential market for business in the Southeast Asia Region. The modern retail industry for the Fast-Moving Consumer Goods (FMCG) category in Indonesia

is growing rapidly, where the highest recorded growth occurred in the mini market and supermarket segments [14]. so that Indonesia becomes the most potential market for business in the Southeast Asia Region. The modern retail industry for the Fast-Moving Consumer Goods (FMCG) category in Indonesia is growing rapidly, where the highest recorded growth occurred in the mini market and supermarket segments [14]. so that Indonesia becomes the most potential market for business in the Southeast Asia Region. The modern retail industry for the Fast-Moving Consumer Goods (FMCG) category in Indonesia is growing rapidly, where the highest recorded growth occurred in the minimarket and supermarket segments [14].

The retail sector as a provider of goods or services used in everyday life has long started digital transformation in its business. With the use of the right model, digital technology can help the retail industry collect valuable customer data. Currently the use of data is one important aspect for a business in making decisions. This is natural because nowadays business competition is getting tougher. The data can later be used as insight to achieve competitive advantage. Business analytics involves the utilization of data, information technology, statistical analysis, mathematical models, or computer-generated models to gain a deeper understanding of business operations and to make informed decisions based on factual evidence. Business analytics involves converting data into actionable steps or strategies within the framework of making informed decisions and addressing challenges. [3]. Digital transformation in the retail sector is very important in order to survive and be in accordance with the lifestyle of its consumers. Competition in the retail sector is currently getting higher, causing retailers to find ways to provide products that are closest to consumer demand. By understanding what consumers need and being able to meet those needs, it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the amount of

revenue to be earned [11]. Digital transformation in the retail sector is very important in order to survive and be in accordance with the lifestyle of its consumers. Competition in the retail sector is currently getting higher, causing retailers to find ways to provide products that are closest to consumer demand. By understanding what consumers need and being able to meet those needs, it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the amount of revenue to be earned [11]. Digital transformation in the retail sector is very important in order to survive and be in accordance with the lifestyle of its consumers. Competition in the retail sector is currently getting higher, causing retailers to find ways to provide products that are closest to consumer demand. By understanding what consumers need and being able to meet those needs, it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the amount of revenue to be earned [11]. Competition in the retail sector is currently getting higher, causing retailers to find ways to provide products that are closest to consumer demand. By understanding what consumers need and being able to meet those needs, it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the amount of revenue to be earned [11]. Competition in the retail sector is currently getting higher, causing retailers to find ways to provide products that are closest to consumer demand. By understanding what consumers need and being able to meet those needs, it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the

amount of revenue to be earned [11]. it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the amount of revenue to be earned [11]. it will directly increase sales. Consumer needs and buying behavior can be predicted with big data-based technology, which can provide information to retailers to determine the products or services offered, the shopping experience, pricing and the amount of revenue to be earned [11].

The big data trend describes a broad area that includes both business and technology. Big data refers to a kind of technology which involves data that is extremely large in size, constantly changing, and varied in nature, making it challenging for traditional technology, knowledge and infrastructure to manage it efficiently [12]. All supermarkets use a computerized system to store sales and purchase data, so large transaction data will be generated. Collection of transaction data can be processed to produce useful information to increase sales. In addition, transaction data can also be used as an important source of information in maintaining business existence. This makes the use of data mining techniques for the analysis of transaction data necessary. Data mining, which involves finding particular patterns in large amounts of data, is the use of pattern recognition technology along with statistical and mathematical techniques to produce valuable information. [8]. Transaction data from each consumer is considered a shopping basket, where the contents of each consumer's shopping basket will vary. From the transaction data that has been collected, it can be used to find patterns of consumer behavior from the goods purchased by consumers.

The association rule is a tool used in data mining to find correlations or associative patterns between products that customers purchased as a set or list. Market basket analysis, also known as shopping basket analysis, is a specific application of the association rule method used to analyze consumer behavior within a particular group. It involves examining the purchasing patterns of

customers to understand how items are related and frequently bought together [5]. In market basket analysis, the Apriori algorithm is frequently used. To find the item in a database that appears the most frequently, R. Agrawal and R. Srikant created the Apriori algorithm. The fundamental idea behind the Apriori algorithm is to find a frequent item set of a specific length using a collection of frequent patterns of various lengths, and then to compare the pattern occurrence counts with the data gathered from the database. In this case, the Apriori Algorithm will have to repeatedly scan the database, especially if there is a lot of data. A new algorithm has been created with the aim of addressing the limitations of the a priori algorithm. The FP-Growth algorithm is among them. The frequent itemset can be determined with just two scans of the database using the FP-Growth algorithm. The technique employed in the FP-Growth algorithm to identify recurring itemsets involves an expanded implementation of a prefix tree, which is referred to as an FP-Tree. The Apriori and FP-Growth algorithms are two examples of association rule algorithms that work with a horizontally structured database format. Equivalence Class Transformation, commonly known as ECLAT, is a technique for association rules that employs a vertical data layout. The Equivalence Class Transformation algorithm is principally focused on discovering common elements within a set of transactions and is solely designed to operate on a database with vertical orientation.

In research [13] describes and reviews the basis of association rules mining and compares three pattern mining algorithms (Apriori, FP-Growth and ECLAT) which concludes that ECLAT and FP-Growth are better than Apriori in terms of execution time and memory usage. Research [7], FP-Growth and ECLAT have almost the same performance, but FP-Growth is slightly better than ECLAT, so we recommend FP-Growth more. Research [4], concluded that based on time efficiency, FP-Growth requires less time to generate frequent itemsets in a small datasets. In terms of larger datasets, the Apriori algorithm is found to be faster in generating frequent item sets. However, the ECLAT

algorithm performs equally well in both small and large datasets, taking a comparable amount of time to generate frequent item sets. Based on these findings, the paper concludes that the ECLAT algorithm exhibits superior performance and time efficiency compared to the other two algorithms.

Based on the description provided, the study to be discussed in this study is to compare the three methods of association rules in frequent pattern mining, namely the Apriori algorithm represents the candidate generation type, FP-Growth represents the pattern growth type and ECLAT represents the vertical format in the use of large data. In addition, this research will first analyze the complexity of the algorithm based on time complexity, namely the time required to run the algorithm. These three methods will be applied to supermarket transaction data with a total of 136,202 transactions with 1,825,632 data rows. In addition, this study will discuss in more depth the managerial implications that can be generated based on the association rules obtained.

## LITERATURE REVIEW

### Data Mining

Data mining involves the identification of significant patterns and trends within vast quantities of data. It is a process that aims to extract valuable insights and knowledge from large datasets [8]. Data mining is a rapidly expanding field with a reputation for being extremely powerful. Businesses make up the majority of users who use data mining due to their regular collection of large amounts of data and a desire to understand it. This is done in an effort to strengthen their comparative advantage while boosting profits [2]. Association rule mining is one of multiple data mining techniques available for extracting valuable information.

### Association Rule Mining

Association rule mining is a method in data mining that uses the relationship of an item with other items contained in an item set to predict patterns from a data set [10]. The purpose of the association rule is to find the joint value of the variable  $\mathbf{x} = (x_1, x_2, \dots, x_r)$  that often appear together

in a database. Association rules are usually applied to binary data  $x_j \in \{0,1\}$ , which is usually used in market basket analysis. The database structure design of the association rule mining algorithm is based on horizontal and vertical data formats. Illustrations of the database structure are presented in Tables 1 and 2.

Table 1. Illustration of horizontal data format

TID	Items
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3

Table 2. Illustration of vertical data format

Items	TID set
I1	{T100, T400, T500}
I2	{T100, T200, T300, T400}
I3	{T300, T500}
I4	{T200, T400}
I5	{T100}

For example,  $\mathbf{i} = \{i_1, i_2, \dots, i_m\}$  is an item set.  $D$  is a database consisting of a collection of transactions, where  $T \subseteq \mathbf{i}$ . Each transaction is associated with an identifier i.e., a transaction ID called a TID. For example,  $A$  is the product that the consumers purchases and  $B$  is the other product that the consumer purchases. A transaction  $T$  said to be contain  $A$  If  $A \subseteq T$ . An association rule is an implication of form  $A \Rightarrow B$ , Where  $A \subset I, B \subset I, A \neq \emptyset$  And  $A \cap B = \emptyset$ . rules  $A \Rightarrow B$  applies to transaction sets  $D$  with support  $s$ , Where  $s$  is the percentage of the transaction  $D$  which contain  $A \cup B$ . This is considered an opportunity  $P(A \cup B)$ . rules  $A \Rightarrow B$  have confidence  $c$  in transaction sets  $D$ , Where  $c$  is the percentage of the transaction  $D$  which contain  $A$  and also contain  $B$  or it can be written as a conditional probability  $P(B | A)$ . In short it can be written:

$$\text{support}(A \Rightarrow B) = P(A \cup B) = \frac{\text{The number of transactions containing } A \text{ and } B}{\text{Transaction totals}} \quad (1)$$

$$\text{Confidence}(A \Rightarrow B) = P(B | A) = \frac{\text{The number of transactions containing } A \text{ and } B}{\text{Number transaction containing } A} \quad (2)$$

Minimum support and minimum confidence are the threshold values used to determine the best association rules. Support and confidence values are between 0-100%.

### Market Basket Analysis

Market basket analysis is an intuitive way of applying association rules to analyze the purchasing behaviors of customers, whereby it identifies correlations between the products placed in their shopping carts [2]. Market basket analysis examines the patterns of customers' purchasing behavior to determine which products have a tendency to be purchased together. This analysis is used to identify opportunities for cross-promotion or bundling of products, where related items can be marketed together as a package [1].

Market basket analysis utilizes Point of Sale (POS) customer transaction data as input to extract valuable product associations. It identifies relationships between products based on transaction patterns, such as the likelihood of customers purchasing both product A and product B together [1]. This analysis automates the process of discovering which items are frequently bought together. Retailers find market basket analysis particularly popular as it helps them make sense of customer transactions and uncover valuable insights, especially considering their extensive product offerings.

The several methods in market basket analysis include: the Apriori, FP-Growth and the ECLAT algorithm. The main difference between the three algorithms is the efficiency of the algorithm on a given data set.

### Apriori Algorithm

The Apriori algorithm was first introduced by R. Agrawal and R. Srikant to find the highest frequent from a database. The Apriori algorithm is used to search for frequent items/item sets in transactional databases. In [6], the Apriori algorithm consists of two stages of the process as follows:

#### 1. Join step

In the a priori algorithm,  $L_{k-1}$  used to find

$L_k$  For  $k \geq 2$ .  $L_k$  is a set of k-itemset that

meets the minimum support value. To determine  $L_k$ , a set of candidates  $k$ -itemset generated by combining  $L_{k-1}$  with himself. This set of candidates is generated as  $C_k$ , Where  $C_k$  is a set of candidates k-itemset.

#### 2. Pruning step

$C_k$  is a superset of  $L_k$ , where each member can be frequent or infrequent, but for all frequent  $k$ -itemset included in  $C_k$ . The process of scanning the database to determine the number of occurrences of each candidate in  $C_k$  will make a determination  $L_k$ . For example, all candidates who have a support count of more than or equal to the predetermined minimum support are called frequent and therefore are included in the list  $L_k$ . There is a possibility  $C_k$  contains a very large number of candidates, so it can involve very heavy computations. To reduce the number of candidates  $C_k$  then the A priori property is used. All  $(k-1)$ -itemset which is infrequent cannot be a subset of a frequent k-itemset.

### FP-Growth Algorithm

The FP-Growth algorithm offers an alternative approach to identify frequent item sets in a dataset. Unlike the Apriori algorithm, which requires multiple scans of the database, FP-Growth only needs two scans. This efficiency is achieved by constructing an FP-Tree, a compact data structure that enables efficient mining of frequent item sets. With FP-Growth, the most frequently occurring data set or frequent item set can be determined more quickly, especially when dealing with large datasets.

The FP-Growth algorithm employs a data structure known as the FP-Tree, which is a prefix tree commonly used for finding frequent item sets. With the FP-Tree, the algorithm directly extracts frequent item sets without the need for

intermediate candidate generation. The process involves two stages: first, constructing the FP-Tree using a divide and conquer approach, and then implementing the FP-Growth algorithm to identify the frequent item sets. This two-step process enables efficient selection of frequent item sets from the FP-Tree, providing a streamlined approach for mining associations in the data.

### **ECLAT Algorithm (Equivalence Class Transformation)**

The Apriori and FP-Growth algorithms are examples of association rules methods with horizontal data formats. One method of association rules that uses a vertical data format is ECLAT or Equivalence Class Transformation. The main task of the Equivalence Class Transformation algorithm is to find frequent items in a transaction group and only works on a vertical layout database. The ECLAT algorithm scans the database to find the frequent item set only once, different from the Apriori algorithm, it takes more time to find frequent item sets because it needs to scan the database repeatedly so that the process takes more time [7]. The ECLAT algorithm finds elements from bottom like depth first search. This algorithm only calculates the support value, while the confidence value is not calculated in this algorithm [7].

The ECLAT algorithm utilizes the concept of equivalence classes to partition the search space into distinct and non-overlapping subspaces. It classifies item sets with the same prefix into the same equivalence class, enabling candidate item set generation to occur exclusively within each class. This approach enhances the efficiency of generating candidate item sets and reduces memory consumption. By leveraging the equivalence layer technology, ECLAT optimizes the process of mining frequent item sets, allowing for more efficient and resource-friendly analysis of large datasets [9].

In summary, the association rule step using the ECLAT algorithm is to transform horizontal format transactional data into vertical format by scanning the dataset once. The support count value of the item set is the length of the TID\_set. Starts with  $k = 1$ , the frequent  $k$ -itemset can be

used to build candidates  $(k + 1)$ -itemset based on A priori property. Calculations are performed with the intersection of the TID\_set of the frequent  $k$ -itemset to calculate the TID\_set of  $(k + 1)$ - the appropriate item set. This process is repeated with  $k$  increasing by 1 each time, until no frequent item set, or candidate item set can be found [6].

### **Algorithm Complexity**

An algorithm is a sequence of processes that are run in sequence to solve a problem. Algorithm complexity is measured by the performance of computing the execution of the algorithm. Algorithm complexity can be divided into two parts: time complexity and space complexity. The time complexity is the time it takes to execute the algorithm and the space complexity is the memory used to execute the algorithm.

### **Retail Business**

Retail is defined as all activities involved in selling goods or services directly to end users for personal, non-commercial use. Thus, retail is the last activity of the distribution channel that connects producers with consumers. Retail is a partner of an agent or distributor. Business has a profit orientation that continues to grow over time. However, businesses also need loyal customers so that business continuity does not stop at one point. Therefore, the ability to meet consumer needs is an important key so that sales can grow. By understanding what consumers need and being able to meet those needs, it will directly increase sales.

## **RESEARCH METHODOLOGY**

### **Data source**

The data used in this study is secondary data, namely transaction data for Supermarket "X" in Central Java in 2021. The data used is 136,202. transaction ID with a total number of data lines of 1,825,632.

### **Research variable**

The raw data obtained from Supermarket "X" is data in Excel form which contains recaps

of daily transactions, where the variables in the data include:

- a. Vendor, namely the name of the distributor or supplier of the goods.
- b. Inventory code, which is a universal product code that follows established rules.
- c. PLU (Price Look Up Unit), namely item identification number used for internal recording.
- d. Inventory name, namely the name of the article or item of an item.
- e. Transaction number, which is a unique code assigned to each transaction.
- f. Sales date is the date of the transaction.
- g. Price, namely the price of each item.
- h. Quantity, the number of items purchased from each transaction.

The variables used in this study are independent variables, namely the inventory name or item name for each transaction and the transaction number. In this study, it did not pay attention to the quantity of goods purchased and also the price of these goods.

### Analysis Step

The analytical steps performed in this study are as follows:

1. Perform data preprocessing.
  - a. Conducting literature studies related to market basket analysis using the Apriori, FP-Growth and ECLAT methods.
  - b. Collect secondary data on Supermarket "X" transactions in 2021. The raw data used in this study is in the form of Excel with a total of 136,202 transactions.
  - c. Read transaction data and retrieve the required information, namely the inventory name or the name of the item purchased and the transaction number.
  - d. Delete transaction numbers that contain items that have been canceled or refunded.
2. Comparing Apriori, FP-Growth and ECLAT algorithms
  - a. Determine the minimum support and confidence values for the algorithms Apriori, FP-Growth, and ECLAT.
  - b. Comparing the time complexity of the Apriori, FP-Growth and ECLAT algorithms.

- c. Analyzing the association rules of the best algorithm obtained based on time complexity.

- d. Sort rules that meet the minimum confidence value based on:
  - Rules with a greater confidence value will rank higher.
  - If there are rules that have the same confidence value, the support value will be seen. Larger support values will rank higher.

3. Analyze consumer purchasing patterns and make recommendations based on market basket analysis results
  - a. Conducting literature studies related to the use of market basket analysis in the retail world.
  - b. Based on step 2 point d, rules that have a high confidence value will be used as a consumer buying pattern.

Develop recommendations for the utilization of the results of analysis of consumer buying patterns as managerial implications.

## DISCUSSION

### Descriptive Analysis

Figure 1. is an example of an item purchased in a shopping basket. One transaction number is a unique shopping cart. In this study there were 136,202 unique shopping baskets. From this entire shopping basket, it can be identified 10 product items that have a high frequency of purchases. The sequence is as shown in Figure 2

<pre> \$ 6363186301202112251940050084 [1] "OHAMN 8F EXT NSQ" [3] "DELMONTE PCH 100 EXT HOT" [5] "DORANI SHP 100 ENGINE" [7] "KOPI TOP GULA AREN SACH 20GR" [9] "NUVO SOAP FML BIRU 602549" [11] "NUVO SOAP FML KNG 602570" [13] "NUVO SOAP FML PINK" [15] "PINKOPI MOCCACHINO BAG" [17] "ROYCO PDS BEEF HP(6*9GR)" [19] "SEDAAP MIE GORENG 200301" [21] "TONG TJI CELUP BLACK TEA HARUM"                 </pre>	<pre> "CHERI TWIST 75 ROASTEDCORN" "DELMONTE TOMATO 10 SACH" "KOPI TOP BARISTA SPEC BLEND" "NABATI SIIP NET FIRE NINGGG 65" "NUVO SOAP FML H3 602593" "NUVO SOAP FML MERAH 602556" "PAKET MONG CC PLOING 120 MANGGA 2+1" "REJOICE SHP SACH(5*2)AN FRIZ" "SEDAAP MIE AYAM BAKANG 200318" "SHENZHI SCRUB 100 REGULAR" "USAGI PUFF GOJR CARAMEL"                 </pre>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 1. Example of items purchased in 1 shopping basket

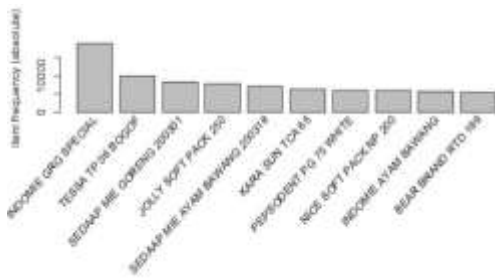


Figure 2. Top 10 most frequent items

Figure 2 is an item that has a purchase frequency with the 10 highest orders. The details are in Figure 3.

Item	Frequency
INDOMIE GRG SPECIAL	18690
TESSA TP 06 BOGOF	9880
SEDAAP MIE GORENG 200301	8260
JOLLY SOFT PACK 250	7790
SEDAAP MIE AYAM BAWANG 200318	7036
KARA SUN TCA 65	6251
PEPSODENT PG 75 WHITE	5951
NICE SOFT PACK NP 200	5838
INDOMIE AYAM BAWANG	5651
BEAR BRAND RTD 189	5421

Figure 3. Most frequent items

Items that have the highest frequency of purchases are Indomie goreng special, where out of 136,202 shopping baskets, 18,690 shopping baskets have Indomie goreng special. The second place is Tessa tissue buy one get one free, where there are 9,880 shopping baskets that buy Tessa tissue, and so on can be seen in figure 3. There are 4 items of instant noodles that are included in the top 10 frequency with the highest purchases, which means instant noodles are goods favorites purchased by consumers.

### Algorithm Complexity Analysis

Time complexity is the time it takes to execute an algorithm. A comparative analysis of the running time between his three algorithms, Apriori, FP-Growth and ECLAT, shows that the ECLAT algorithm is the shortest algorithm with an average running time of 0.788. Therefore it can be concluded that the ECLAT algorithm is the most effective algorithm when compared to the Apriori and FP-Growth algorithms.

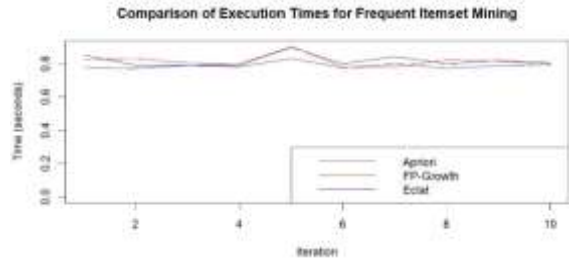


Figure 4. Comparison of execution times

Algorithm	Mean_Time
Apriori	0.8182324 secs
FP-Growth	0.8157283 secs
Eclat	0.7880944 secs

Figure 5. Mean Time of 3 algorithms

The space complexity is the memory used to execute the algorithm. A comparative analysis of the running memory between his three algorithms, Apriori, FP-Growth and ECLAT, shows that the ECLAT algorithm is the smallest algorithm with a mean memory usage of 1.03579 MB. Therefore it can be concluded that the ECLAT algorithm is the most effective algorithm when compared to the Apriori and FP-Growth algorithms. After obtaining the most effective algorithm of time complexity and space complexity, then an analysis of the formation of association rules is carried out using ECLAT.

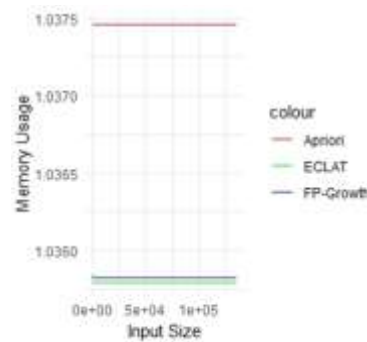


Figure 6. Comparison of memory usage

Algorithm	Mean_Memory_Usage_MB
Apriori	1.03746
FP-Growth	1.03582
ECLAT	1.03579

Figure 7. Mean memory usage of 3 algorithm.



### Formation of Association Rules

The minimum support used in the formation of this association rule is 1%, or at least the item has been purchased 1,362 times, which can be categorized as having a strong association. In the research results, the highest support value is the consumer transaction buying Indomie goreng special and Indomie ayam bawang together. Out of a total of 136,202 transactions, there were 3,698 shopping baskets that bought Indomie goreng special and Indomie ayam bawang together, with a support value of 0.0271, or 2.71%.

The entire association rule generated from the transaction data can be viewed in Figure 6. Based on the obtained association rule, it can be seen that the items purchased together are mainly instant noodles with multiple combinations. Different brands and flavors. It can be said that instant noodles are a very popular item that customers buy together.

ITEMS	SUPPORT	COUNT
[1] [INDOMIE GRG SPECIAL, INDOMIE SOTO MRE]	0.01117458	1522
[2] [TEM GOREK K, TEH NYAMU K]	0.01058722	1442
[3] [GULAKU PREMIUM PUTIH 1 KG, INDOMIE GRG SPECIAL]	0.01048443	1428
[4] [RDR SCH SCH 4D CHOCO, RDR SCH SCH 4D KENTAL MANIS]	0.01651224	2249
[5] [INDOMIE GRG SPECIAL, SUPERMI AYAM BAWANG]	0.01946374	2651
[6] [RDR SCH SCH 4D KENTAL MANIS, INDOMIE GRG SPECIAL]	0.01008062	1373
[7] [INDOMIE GRG SPECIAL, SEDAAP MIE SOTO 200325]	0.01189410	1620
[8] [SEDAAP MIE GORENG 200321, SEDAAP MIE SOTO 20032E]	0.01433900	1953
[9] [SEDAAP MIE AYAM BAWANG 200318, SEDAAP MIE SOTO 20032E]	0.01408937	1919
[10] [INDOMIE AYAM SPECIAL, INDOMIE GRG SPECIAL]	0.02251892	3067
[11] [INDOMIE AYAM BAWANG, INDOMIE AYAM SPECIAL]	0.01066798	1453
[12] [INDOMIE GRG SPECIAL, NICE SOFT PACK NP 200]	0.01081482	1473
[13] [INDOMIE AYAM BAWANG, INDOMIE GRG SPECIAL]	0.02715085	3698
[14] [INDOMIE GRG SPECIAL, JOLLY SOFT PACK 250]	0.02201157	2936
[15] [INDOMIE GRG SPECIAL, KARA SUN TCA 83]	0.01318830	1758
[16] [INDOMIE GRG SPECIAL, SEDAAP MIE AYAM BAWANG 200318]	0.01394960	1872
[17] [SEDAAP MIE AYAM BAWANG 200318, SEDAAP MIE GORENG 200301]	0.02159293	2941
[18] [INDOMIE GRG SPECIAL, SEDAAP MIE GORENG 200301]	0.01369290	1855
[19] [INDOMIE GRG SPECIAL, TESSA TP 06 8000F]	0.01719505	2342

Figure 6. The results of association rules with a support value of 1%.

### CONCLUSION

Collection of transaction data can be processed to produce useful information. Based on the results of the analysis of the execution time required by the algorithm, the results show that the ECLAT algorithm is the most effective algorithm because it has the fastest execution time and smallest memory usage. The support value used in forming the association is 1%, with 19 rules being produced. From the results of these rules, the highest support value was generated by the purchase of Indomie goreng special and Indomie ayam bawang, where as many as 1,362 shopping transactions bought these two items together or 2.71% of the total transactions. By knowing the results of the rules obtained, that the dominance of consumer purchases is instant

noodle products, so that instant noodles can be used as cross-selling promotional products with products that have low purchases, so that products with low purchase rates can be supported by instant noodles as an attraction. In addition, to be able to have a more personal relationship with consumers, they can create a membership program, so that the items purchased by consumers, supermarket entrepreneurs can know the consumer's profile more personally, so that the promos carried out can be right on target according to consumer shopping habits.

### REFERENCES

- [1] Blattberg, RC, Do Kim, B., and Neslin, SA (2008), Database Marketing: Analyzing and Managing Customers, Springer Science+Business Media, New York
- [2] Cios, KJ, Pedrycz, W., Roman WS, and Lukasz AK (2007), Data Mining: A Knowledge Discovery Approach, Springer Science+Business Media, New York.
- [3] Evans, J. (2017), Business Analytics. Pearson Education Limited, England
- [4] Gayathri, GS (2017), "Performance Comparison of Apriori, ECLAT and FP-Growth Algorithm for Association Rule Learning", International Journal of Computer Science and Mobile Computing, Vol. 6 Issue 2, p. 81-89
- [5] Gunadi, G. and Dana, IS (2012), "The Application of the Data Mining Market Basket Analysis Method to Book Product Sales Data Using the Apriori Algorithm and Frequent Pattern Growth (FP-Growth): Case Study of Printing PT. Gramedia", MKOM Telematics Journal, Vol. 4 No. 1, pp. 118-132
- [6] Han, J., Micheline, K., and Pei, J. (2012), Data Mining Concepts and Techniques 3rd edition, Elsevier Inc., USA
- [7] Heaton, J. (2016), Comparing dataset characteristics that favor the Apriori, ECLAT or FP-Growth frequent itemset mining algorithms. In Southeast Con 2016 (pp. 1-7). IEEE.
- [8] Larose, DT and Larose, CD (2014), Discovering Knowledge in Data. 2nd edition, John Wiley and Sons, Inc., New Jersey
- [9] Ma, Z., Juncheng, Y., Taixia Z., and Fan, L (2016), "An Improved ECLAT Algorithm form Mining Association Rules Based on Increased Search Strategy", International Journal

*of Database Theory and Application*, Vol. 9  
No. 5, pp. 251-266

- [10] Olson, D. L and Delen, D. (2008), *Advanced Data Mining Techniques*, Springer, Heidelberg
- [11] Ong, JO, Sutawijaya, AH, and Ahmad, BS (2020), "Innovation Strategy for Modern Retail Business Models in the Industrial Age 4.0", *Scientific Journal of Business Management*, Vol. 6, No. 02
- [12] Pujianto, A., Awin, M., and Rachmawati, N. (2018), *Utilization of Big Data and Consumer Privacy Protection in the Digital Economy Era*, Vol. 15 No. 2, pp. 127-137
- [13] Sharma, A. and Anita, G. (2021), "Association Rule Mining Algorithm: A Comparative Review", *International Research Journal of Engineering and Technology (IRJET)*, Vol. 03 issue 11, p. 848-853
- [14] Soliha, E. (2008), *Analysis of the Retail Industry in Indonesia*, *Journal of Business and Economics*, Vol. 15, No. 2, pp. 128-142