Bit-Tech

Vol.2, No.3, Mei 2020
Available online at: http://jurnal.kdi.or.id/index.php/bt

Analysis of the Application of Customer Purchase Mining Data on Paint Sales Using Apriori Algorithm (Case Study: PT Indowarna Cemerlang Indonesia)

Maman Novian[1)*], Rino[2)],

[1)2)] *Buddhi Dharma University*
Jl. Imam Bonjol No 41 Karawaci Ilir, Tangerang, Indonesia
[1)] maman_novian@yahoo.com
[2)]rino@ubd.ac.id

Article history:

Received 16 March 2020;
Revised 3 April 2020;
Accepted 8 April 2020;
Available online 30 Mei 2020

Keywords:
Customer Purchase Patterns
Data Mining
Apriori Algorithm
Cat

Abstract

Sales transaction data is one thing that can be used for making business decisions. Most sales transaction data is not reused, and is only stored as an archive and only used for making a sales report. Paint sales data is one science that can be applied in cases like this. Sales transactions that are not utilized properly can be extracted and reprocessed into useful information using data mining techniques. Using one of the data mining methods, namely the a priori algorithm, sales transaction data can be reprocessed so that it can produce a consumer buying pattern. This consumer buying pattern will later help companies make business decisions. PT Indowarna Cemerlang Indonesia is a company engaged in the paint trade, where the main activity is selling various wall paints, oil / wood paints, NC paints (car paints), epoxy paints (floor paints), depo-proof (anti leaked). PT Indowarna Cemerlang Indonesia does not reuse sales transaction data resulting from its sales activities. This data is only used as a reference for making sales reports and as an archive only, causing accumulation of data and unknown paint brands that are often sold or those that are of interest to customers. Therefore, the author takes the title application of data mining analysis of customer purchase patterns in paint sales using a priori algorithm. By doing this research, it is expected to provide results in the form of information that can be useful for related parties and can design sales strategies to increase company turnover.

I. INTRODUCTION

Sales transaction data is one thing that can be used for making business decisions. Most sales transaction data is not reused, and is only stored as an archive and only used for making a sales report. Paint sales data is one science that can be applied in cases like this. Sales transactions that are not utilized properly can be extracted and reprocessed into useful information using data mining techniques. Using one of the data mining methods, namely the a priori algorithm, sales transaction data can be reprocessed so that it can produce a consumer buying pattern. This consumer purchasing pattern will help companies to make business decisions [1]. PT Indowarna Cemerlang Indonesia is a company engaged in the paint trade, where the main activity is selling various wall paints, oil / wood paints, NC paints (car paints), epoxy paints (floor paints), depo-proof (anti leaked). PT Indowarna Cemerlang Indonesia does not reuse sales transaction data resulting from its sales activities. This data is only used as a reference for making sales reports as well as being an archive [1], causing accumulation of data and unknown paint brands that are often sold or those that are of interest to customers [2]. The analysis of previous related research becomes one of the references of this research, namely research that has been conducted by Gunadi G & Sensuse D conducted research using data mining methods with a priori algorithm and FP Growth in determining effective sales and marketing strategies at PT Gramedia. Where the priori algorithm and FP Growth are the algorithms used to help find a number of association rules from the book product sales transaction database. In the association rules, the FP growth algorithm found a relationship between the itemset and the strength level of the support and confidence association rules of 0.273, while the a priori algorithm association rules found a relationship between itemset with the strength level of the support and

confidence association rules of 0.725. So it is concluded that the a priori algorithm can help in determining the sales and marketing strategy of books [2]. The a priori algorithm that works by searching for and finding associated patterns (association rule) among products marketed, [4] the association rule which This is done through a mechanism to calculate the support and confidence of an item relationship. An association rule is said to be interesting if the value of support is greater than the minimum support and also the value of confidence is greater than the minimum confidence. This a priori algorithm will be suitable to be applied if there are several relationship items to be analyzed. One of them that can be applied is in the business sector in determining the pattern of buying paint [3]. Based on the above problems, this study aims to create a web-based application to process sales transaction data so that it knows which products are interconnected and can produce information about consumer purchasing patterns which will be used to help owners make business decisions. [1] in this study the authors decided to use the a priori algorithm because "Apriori algorithm works well with large databases but FP Growth algorithm works badly with large databases" according to [5] the a priori algorithm works well with databases than FP-Growth algorithm works poorly with huge database. The application of the a priori algorithm in this study is expected to find patterns in the form of products that are often purchased simultaneously. This pattern can be used to design sales strategies and place frequently purchased products together in an area that is close to each other [3]. The company at least knows better how they should increase the paint stock for the company. [4] By carrying out this research, it is hoped that it can provide results in the form of information that can be useful for related parties. [3] And can design sales strategies to increase company turnover. Therefore the author takes the title "Application of Data Mining Analysis of Customer Purchase Patterns in Paint Sales Using the Apriori Algorithm (Case Study: PT Indowarna Cemerlang Indonesia)".

II. METHODS

1.1. Apriori Algorithm

[6] a priori is one of the most well-known algorithms in data mining in finding data patterns or data occurrence / frequency patterns. The usual a priori algorithm is used to find customer buying patterns at a minimarket based on purchase transactions. In a priori algorithm, the term support value is the value used to measure the appearance of certain data compared to the total data. [7] The a priori algorithm is an algorithm that performs frequent itemset searches using the association rule technique [1]. To find the association rule from a data set, the first step that must be done is to find the frequent itemset first. Frequent itemset is a set of items that often appear simultaneously . The importance of an association can be determined by two benchmarks, namely support and confidence. Support is the value of support or the percentage of the combination of an item in the database, while confidence is the value of certainty, which is the strong relationship between items in the association rule. The main process carried out in the a priori algorithm to get frequent itemset is:

1. Join (merging). This process is done by combining items with other items so that no more combinations can be formed.

2.Prune (pruning). The trimming process is the result of a combined item then trimmed using a predetermined minimum support. [3] a priori algorithm is a data collection algorithm with an associative rule (association rule) to determine the associative relationship of a combination of items.

1. Analysis of High Frequency Patterns with the Apriori Algorithm Looking for combinations of items that meet the minimum requirements of the support values in the database. The support value of an item is obtained using the following formula:

The Support value of 1 item is obtained using the formula :

$$\text{Support (A)} = \frac{\text{(Number of Transactions Containing A) /}}{\text{(Total Transactions)}}$$

The Support value of 2 items is obtained using the formula:

Support (A, B) = P (A∩B)

$$\text{Support (A, B)} = \frac{(\sum \text{Transactions Containing A and B})}{(\sum \text{Transaction})}$$

2. Establishment of Association Rules
After all high frequency patterns are found, then look for association rules that meet the minimum requirements for confidence by calculating the confidence of the associative rule A U B. The confidence value of the A U B rule is obtained by the following formula:

$$\text{Confidence} = (A \Rightarrow B) = P(B / A) = \frac{(\sum \text{Transactions Containing A and B})}{(\sum \text{Transactions Containing A})}$$

To determine which association rule to choose, it must be sorted based on support × confidence. N rules are taken that have the greatest payoff.

1.2. Association Rules

[8] association rule is a method in machine learning or data mining that is used to determine the relationship or relationship between one item and another in a database. The function in association rules is often referred to as market basket analysis which aims to find correlations between sets in certain items. This association rule can also be applied to databases with large dimensions consisting of transaction data. [9] Association analysis or association rule mining is a data mining technique to find associative rules between a combination of items. Association analysis is also known as a data mining technique which is the basis for another data mining technique. In particular, one of the stages of association analysis that has attracted the attention of many researchers is to produce efficient algorithms, namely high frequency pattern analysis (frequent pattern mining).

[10] association rule mining is a mining technique to find associative rules between a combination of attributes. Association rules will find certain patterns that associate data with one another. If it is assumed that the goods sold at the supermarket are all, then each item will have a variable value that will indicate whether the item is there or not in one transaction or one shopping basket. Bolean patterns that can be used to analyze items that are often purchased simultaneously. The pattern can be formulated in an association rule.
For example, consumers will usually buy coffee & milk which is shown as follows:

Coffee → milk (support = 2%, confidence = 60%)

Association rule requires a size variable that is determined by the user to determine the extent or amount of output that the user wants. Support & confidence is a measure of the trust & usefulness of a pattern that has been found. A support value of 2% indicates that the total number of consumer transactions buying coffee & milk simultaneously is 2%. Meanwhile, 60% confidence indicates that consumers buy coffee and definitely buy milk by 60%.

1.3. CRISP-DM (cross-industry standard process for data mining)

[11] cross-industry standard process for data mining (CRISP-DM) which was developed in 1996 by analysts from several industries such as Daimler Chrysler, SPSS, and NCR. CRISP DM provides a standard data mining process as a general problem-solving strategy of the business or research unit. In CRISP-DM, a data mining project has a life cycle that is divided into six phases. All sequential phases that exist are adaptive. The next phase in the sequence depends on the output of the previous phase. The important relationships between the phases are depicted by arrows. For example, if the process is in the modeling phase. Based on the behavior and characteristics of the model, the process may have to return to the data preparation phase to further refine the data or move forward to the evaluation phase.

[12] CRISP-DM a data mining project has a life cycle that is divided into six phases, the six phases are as follows:

1. Business understanding phase.
2. Data understanding phase.
3. Data processing phase (data preparation phase).
4. The modeling phase.
5. Evaluation phase (evaluation phase).
6. Deployment phase.

1.4. Data Mining

[10] Data mining is a series of processes to explore added value in the form of information that has not been known manually from a database. [13] data mining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from large databases. [6] in data mining there are so many algorithms / methods / techniques of extracting or seeking knowledge or information. Each algorithm / method / technique has different functions and purposes. Following are data mining groupings based on function and purpose:

a. Description
Description of the process for finding / approving patterns that often appear and turning these patterns into rules that can be used to simplify an activity. For example in a supermarket, a customer who often buys product A and product B simultaneously and repeatedly, then the supermarket management changes the catalog of goods by completing products A and B in the same or adjacent places, so that customers have no trouble buying back the product. One of the algorithms included in the description is a priori algorithm.

b. Classification
The grouping is based on the relationship between the criterion variable and the target variable. For example, the classification of the impact of an earthquake, namely heavy damage, heavy damage and tsunami, or no impact. The impact of the earthquake was caused by several factors such as the scale, duration, epicenter, distance from the coast and the depth of the earthquake. Algorithms included in the classification are Cart, ID3, C4.5, J48, C5.0, nearest neighboor, naive bayes, and others.

c. Prediction
In general, prediction is almost the same as classification. One of the data mining functions that is often used to predict. The value of the predicted results will be used in the future based on previous data. For example, the price of palm oil in the next 4 (four) months, predictions of visits to tourist attractions in the coming year and so on. The algorithms that are included in the prediction are Rough Set, Cart, ID3, C4.5, J48 and C5.0.

d. Estimate
The definition of an estimate is an estimate / prediction, so that the estimate is almost the same as the classification, the difference lies in the form of grouping, where the estimates are grouped in a numerical direction rather than in a category direction. For example, the estimated income of a sales person selling certain products based on length of service, estimated hotel revenue for a certain month and so on. Algorithms included in the estimation are simple linear regression, multiple linear regression and others.

e. Clustering
Clustering is a grouping of data that has a similarity value (homogeneous). The form of data that can be grouped into clustering is the result of observations, data records, or classes and objects that have similarities. Clustering is different from classification because it does not use a decision / target variable. For example, the grouping of decent and unfit families is given the family hope program (PKH) based on the amount of income, number of dependents, assets, occupation and ownership and condition of the house. Another example is grouping students with academic potential based on national examination scores and university entrance test scores. Algorithms included in the clustering are K-means, K-medoids, K-nearest neighboor, and others.

f. Association

Association is a collection, association, union, or fellowship. In data mining, the association process is a search for attributes that appear / always appear at the same time, such as when product A is purchased, product B is purchased, when product B is purchased, product A is purchased, when product A, B is purchased, product C is purchased, and so on. . The probability of the attribute appearing simultaneously is measured using the confidence value. The algorithm that is included in the association is the association rule.

III. DESIGN

2.1. Business Understanding

This stage, namely understanding the problem of objectives and needs from a business point of view, requires an understanding of the data mining activities to be carried out. In this case, it is the identification of problems that exist in PT Indowarna Cemerlang Indonesia, namely the large number of unorganized paint products between the many sold and interconnected and the imbalance of the stock of goods between the items that are sold and the goods that are interconnected so that the stock turnover of the goods is not controlled. In understanding the problems that exist in the object of research it is very important to be able to determine the initial strategy that is more targeted and can achieve goals. Taking new information is a step that can be done in this research activity. In order to obtain new information, this study uses the sales transaction data of PT Indowarna Cemerlang. Stored transaction data will be extracted again using data mining techniques, for example to determine consumer purchasing patterns in purchasing products. To find out which products are often purchased simultaneously by consumers, it is analyzed by using a shopping basket analysis (market basket analysis). This is an initial strategy that is carried out before later designing an application using the a priori algorithm. With the aim of being able to find product association rules for the purposes of sales strategy in increasing PT Indowarna Cemerlang turnover.

1. Determine Business Goals
The aim is to carry out a sales strategy for paint products, based on the purchasing patterns of consumers.

2. Conduct a situation assessment
This stage requires understanding the objectives and translating them into data mining objectives. The goal is to reuse sales transaction data using data mining techniques with a priori algorithm to look for product purchase patterns that most often appear with association rules in the sale of paint products which can be used to solve sales strategy problems and unbalanced stock of goods between items sold. with interconnected goods, so that the stock turnover of goods is controlled and displays the results of data that have been processed from data mining into an application, as well as organizing paint products that are sold a lot and are interconnected.

2.2. Data Understanding

In this stage, it is necessary to collect initial data that will be used or processed later. The data to be used is product sales transaction data stored in the database that occurred within the last 1 month of March 2020, this data will be used for the analysis of association rules. According to research (John Isharianto, 2018) entitled "the study of the application of data mining market basket analysis to determine sales patterns (item association rules) in Olfish restaurants using a priori algorithm", the data used is transaction data that occurs in just 5 days, because In the a priori algorithm that is examined in each transaction is not how much the number of each item was sold, but what is examined is whether an item was sold or not.

2.3. Preparation Data

This stage includes all activities to build a dataset (data to be processed at the modeling stage) from raw data. From understanding the data in the data understanding phase then the data that has been collected, the author uses the transaction data that has been obtained in the form of an excel file to make it easier for data processing, by removing some unused attributes or cleaning the data.

The data taken here is transaction data when conducting a survey, this study uses data from paint sales transactions at PT Indowarna Cemerlang Indonesia in the last 1 year, starting from March 1, 2019 to February 30,

2020, this research takes the sample data for the last month in March 2020. From the initial data in the form of consumer purchase transaction data to building a dataset, from this dataset which will be applied into a modeling tool, then the data mining process will be carried out. At this stage several steps are carried out in data preparation so that it produces data that can be used in the next process, as follows:

1. Data Cleaning
In the data cleaning, here it removes unused attributes. Data cleaning in general does, among others, fill in missing values, identify outliers, handle noise data, correct inconsistent data, and resolve data redundancy problems due to data integration.

2. Data Reduction
This stage is a step to reduce data dimensions, eliminate fields that are considered not contributing to the final result, and data compression.

3. Data Transformation
In this data transformation stage, the data that has gone through the data cleaning and data reduction stages will then start the transformation process so that the resulting data is quality data, and the formation of attributes or features. data generalization, data normalization, and attribute / feature formation.

2.4. Modeling Phase

At this stage, the modeling phase is where we design a study, conceptualize research starting from the preparation of the data to be used, selecting the data mining technique to be used, and combining the parameters of the data mining technique with optimal values. In this stage, various modeling techniques will be selected and applied and some of the parameters will be adjusted to obtain optimal values. In particular, there are several different techniques that can be applied to the same data mining problem. On the other hand there are modeling techniques which require a special data format. So that at this stage it is still possible to return to the previous stage. Until it comes to the evaluation and ends with the conclusion of the results. This modeling is done using the RapidMiner application with two stages, namely finding the minimum value of support and the minimum value of confidence.

2.5. Evaluation Phase

There is this stage.The evaluation phase is to find the association rules of transaction data that must meet two parameters, namely the minimum value of support and the minimum value of confidence to get good association rules The data set that has been prepared will then be processed using the RapidMiner application, this stage begins preparing the sample data has been determined to be processed and analyzed. The evaluation here will be carried out in depth with the aim of adjusting the results at the modeling stage to match the goals to be achieved in the Business Understanding stage. In the cat sales transaction data, tabular forms or often referred to as association networks will make it easier to analyze and find out how many items are often purchased in each transaction.

2.6. Deployment phase

At this stage, the deployment or deployment phase is the stage of producing data mining activities. The final report regarding the knowledge gained or pattern recognition in the data in the data mining process is easy to understand. In data processing using Rapid Miner, with a sample data of 500 data with a minimum value of Support 3% and a minimum value of 50% Confidence. Generates 2 association rules or rules. And here are the results of the data mining process using Rapid Miner:

Association Rules

1. [DEPPON PAINT GREEN TOSCA-G (506)] -> [DEPPON PAINT CREAM-G (502)] (confidence: 0.551).

2. [DEPPON PAINT SUPER WHITE - P] -> [DEPPON PAINT SUPER WHITE - G)] (confidence: 0.52).

IV. DISCUSSION

3.1. Program View

a. Login page

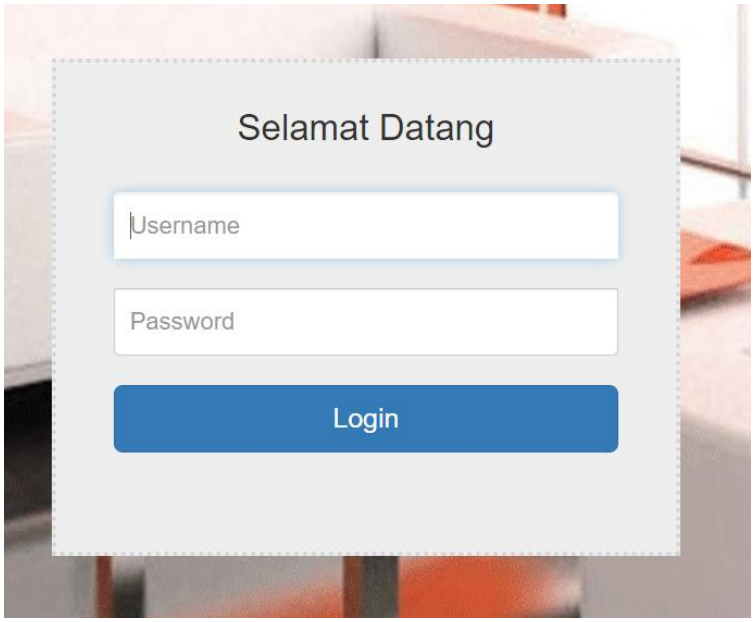

Figure 3.1 Login Page Display

b. Home page

Figure 3.2 Home Page Views

c. Import Data page

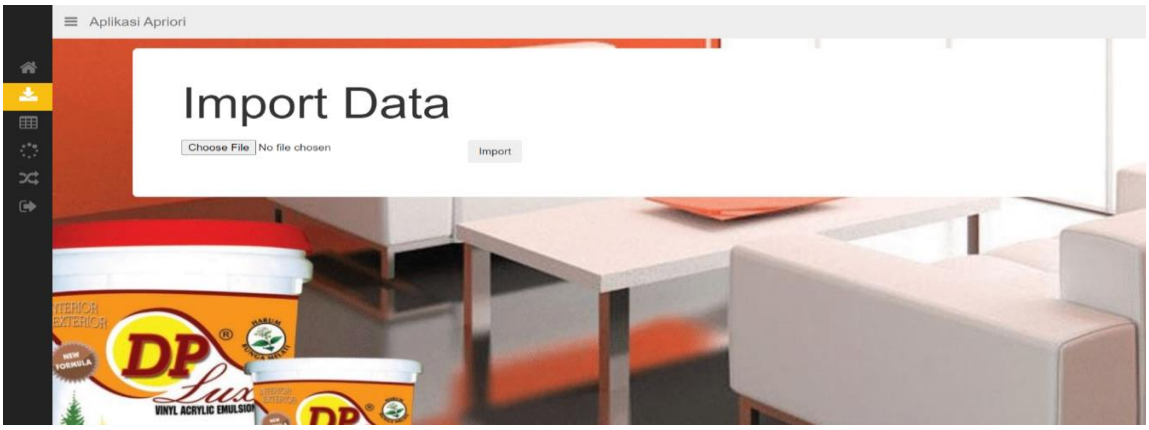

Figure 3.3 Import Data Page Views

d. Sales Transaction Data page

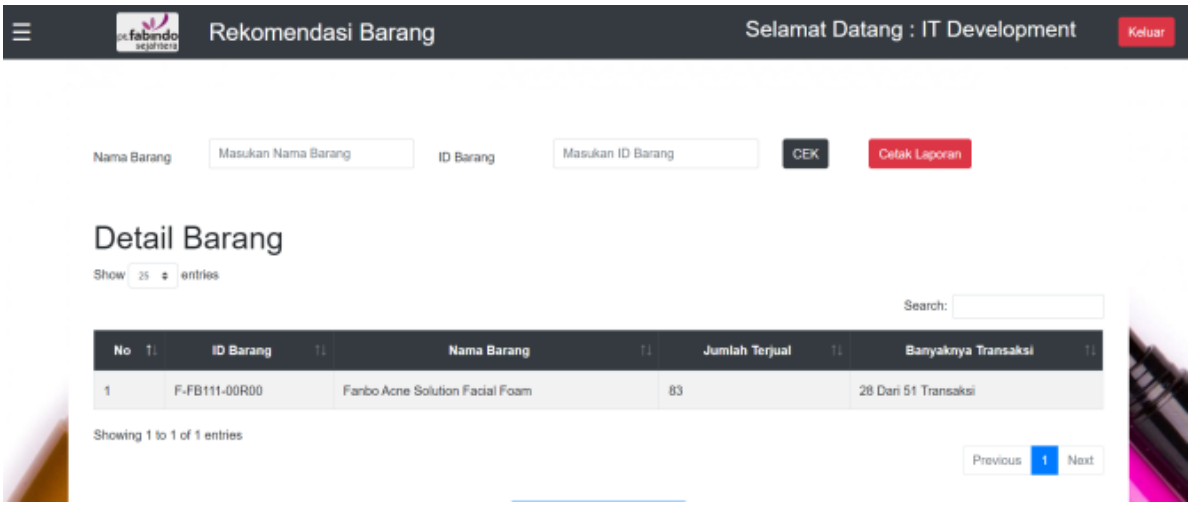

Figure 3.4 Sales Transaction Data Display

e. A priori Process page

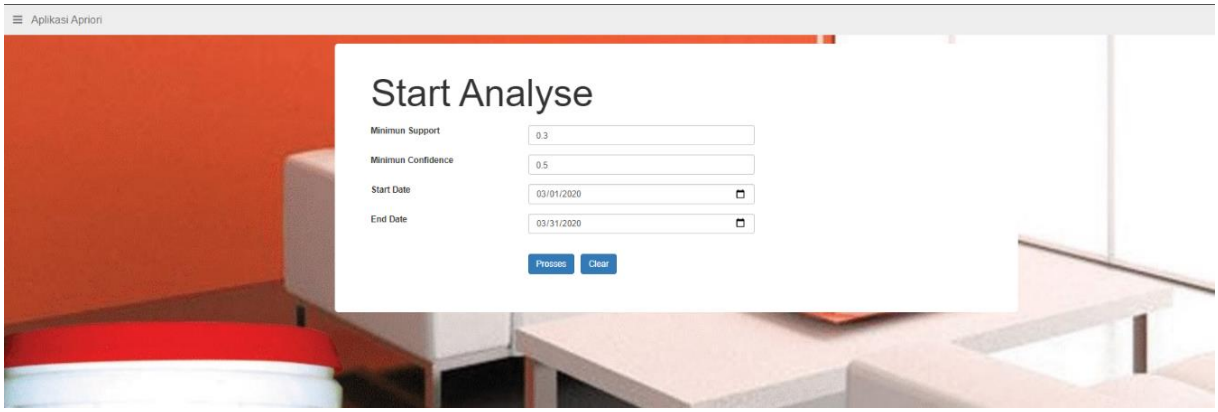

Figure 3.5 Apriori Process Page Views

f. Association Rules Results page

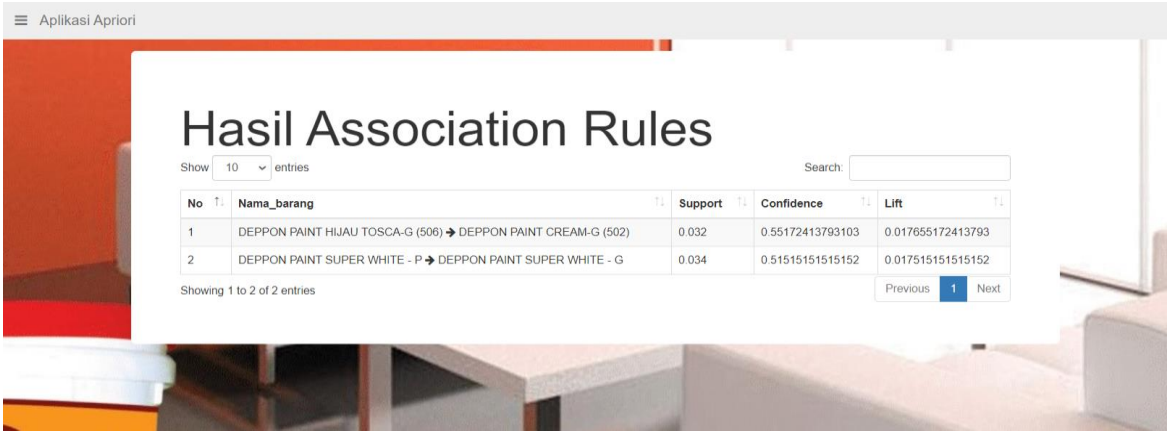

Figure 3.6 Display Association Rules Results

g. Logout page

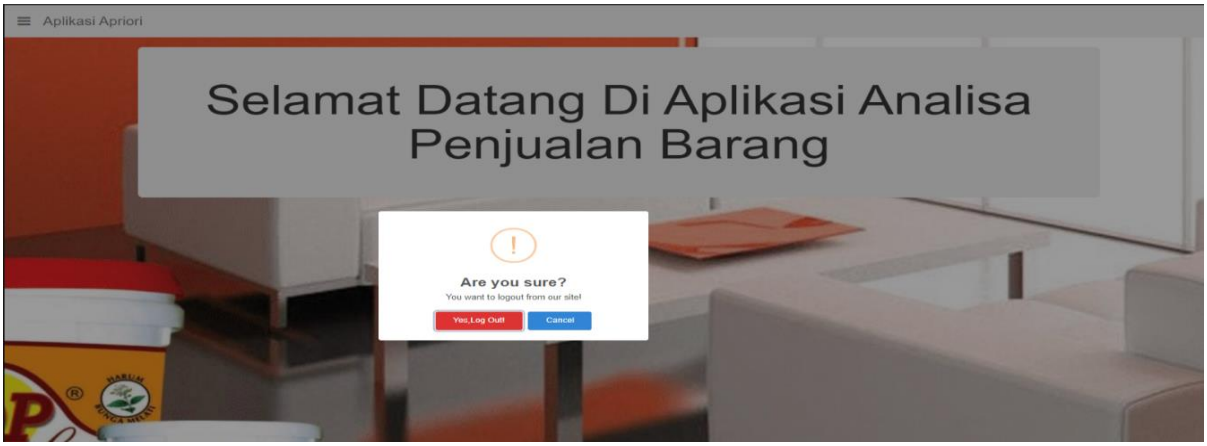

Figure 3.7 Logout Page Views

Figure 3.8 Page Views After Logout

V. CONCLUSIONS

This study can be concluded that the data mining method uses a priori algorithm. By reusing the sales

transaction data which is only an archive at PT Indowarna Cemerlang Indonesia. Based on the analysis and test results, the following conclusions can be obtained:

1. By reusing sales transaction data in data mining techniques using a priori algorithm with support and confidence parameters, we can find customer buying patterns to further increase sales.

2. The information obtained is very useful to assist companies in making decisions and designing sales strategies.

3. With the construction of an application using a priori process, it can produce itemset combination information, so it can present information, namely:

If a customer buys “DEPPON PAINT GREEN TOSCA-G (506)” paint, he will buy “DEPPON PAINT CREAM-G (502)” paint. And if the customer buys "DEPPON PAINT SUPER WHITE - P" then he will buy "DEPPON PAINT SUPER WHITE - G". With a minimum support value of 3% and a minimum confidence value of 50%.

REFERENCES

[1] Y. Kurnia, Y. Isharianto, Y. C. Giap, A. Hermawan, and Riki, “Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm,” *Journal of Physics: Conference Series*, vol. 1175, no. 1, 2019, doi: 10.1088/1742-6596/1175/1/012047.

[1]. Djamaludin, I., & Nursikuwagus, A. (2017). ANALISIS POLA PEMBELIAN KONSUMEN PADA TRANSAKSI PENJUALAN MENGGUNAKAN ALGORITMA APRIORI. 2-8.
[2]. Rulianto, J., & Mustika, W. P. (2019). Implementasi Algoritma Apriori Terhadap Data Penjualan Oli Mobil Berbasis Sistem Pencarian Aturan Asosiasi . 2-11.
[3]. Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. 2-10.
[4]. Ummi, K. (2015). ANALISA DATA MINING DALAM PENJUALAN SPAREPART MOBIL DENGAN MENGGUNAKAN METODE ALGORITMA APRIORI (STUDI KASUS : DI PT. IDK 1 MEDAN). 2-10.
[5]. Kavitha, & Selvi, T. (2016). Comparative Study on Apriori Algorithm and Fp Growth. 3-4.
[6]. Buulolo, E. (2020). Data Mining Untuk Perguruan Tinggi. yogyakarta: Deepublis.
[7]. Santoso, H., Hariyadi, I. P., & Prayitno. (2016). DATA MINING ANALISA POLA PEMBELIAN PRODUK DENGAN MENGGUNAKAN METODE ALGORITMA APRIORI. 2-6.
[8]. Adinugroho, S., & Sari, Y. A. (2018). Implementasi Data Mining Menggunakan Weka. malang: Universitas Brawijaya Press.
[9]. Fauzy, M., W, K. R., & Asror, I. (2016). PENERAPAN METODE ASSOCIATION RULE MENGGUNAKAN ALGORITMA APRIORI PADA SIMULASI PREDIKSI HUJAN WILAYAH KOTA BANDUNG. 2-7.
[10]. Siregar, A. M., & Puspabhuana, A. (2017). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. Jawa Tengah: CV Kekata Group.
[11]. Septiana, Y., & Dharmayanti, D. (2016). PENERAPAN IMPROVED APRIORI PADA APLIKASI DATA MINING DI PERUSAHAAN KALVIN SOCKS PRODUCTION. 2-8.
[12]. Listriani, D., Setyaningrum, A. H., & Eka, F. (2016). PENERAPAN METODE ASOSIASI MENGGUNAKAN ALGORITMA APRIORI PADA APLIKASI ANALISA POLA BELANJA KONSUMEN. 3-8.
[13]. Nofriansyah, D. (2015). *Konsep Data Mining Vs Sistem Pendukung Keputusan*. yogyakarta: Deepublish.