

Association Rules for Product Sales Data Analysis Using The Apriori Algorithm

Jarseno Pamungkas^{1)*}, Yopi Handrianto²⁾,

¹⁾ STMIK Nusa Mandiri, Indonesia ²⁾ Bina Sarana Informatika University, Indonesia

¹⁾ukazseven@gmail.com ²⁾yopi.yph@bsi.ac.id

Submitted : Aug 30, 2020 | **Accepted** : Oct 7, 2020 | **Published** : Oct 7, 2020

Abstract: To increase sales transactions, the company must be able to compete with other competitors so that it requires an appropriate strategy in carrying out the sales process carried out. In addition to the marketing strategy, the company must be able to analyze the products sold based on the number of sales that have occurred so that the company can see which products are more dominant in consumer demand so that the company can determine a more effective sales strategy. PT. Surya Indah City is a company engaged in the sale of various clothing and accessories. In an effort to increase sales of its products, an analysis is needed to be able to increase company revenue by utilizing sales transaction data it has. To analyze the relationship between clothing products and accessories which are more predominantly sold and other available clothing and accessories products, a data mining algorithm is used, namely the a priori algorithm. With the help of the tanagra application to carry out the calculation process, the dominant product that consumers are interested in can be determined. By using two variables that meet support and minimum confidence, it can be concluded that the most sold products are from the type of clothing, namely clothes and pants. It was concluded that the results of the known final association rules, if you buy a shirt, you will buy pants with 50% support and 75% confidence. If you buy pants, you will buy clothes with 50% support and 85% confidence.

Keywords: Apriori Algorithm, Product Sales, Analysis, Tanagra

INTRODUCTION

Clothing and accessories are a very important part of human clothing needs. Various trends and models continue to evolve over time. Nowadays, clothes and accessories are not only a necessity, but have become a will in fulfilling the taste of the wearer (Lestari, 2014). Seeing these conditions, many companies make clothes and accessories as objects of sale to consumers. This business is very promising for every company to get a large enough profit if each company has a good marketing and sales strategy in the face of competition with competitors. PT. Surya Indah City is a company engaged in the sale of clothing and accessories. Many variants of the types of clothing and accessories with various brands are sold at this company, such as various styles of clothes, pants, jackets, sandals, hats, skullcaps, bags, shoes and socks. Along with the increase in product sales, PT. Surya Indah City does not yet have a system that helps to determine the pattern of product purchases by consumers properly because it has not used the sales transaction data it has to be used as a measure in evaluating product purchases that must be done to meet the stock of goods in the warehouse. As a result, the company only predicts conventionally in placing product orders from distributors when the inventory is low (Fauziah & Ratnawati, 2018). The use of sales transaction data can actually be used to help analyze the buying patterns made by consumers, rather than just being used as a data collection (Santoso et al., 2016). In this data set, actually has information that can be used to see the product purchase patterns that occur in each transaction (Putra et al., 2019). Other problems faced by PT. Surya Indah City is the absence of a system that determines the layout of products sold. In fact, the product layout system being marketed will affect product sales because this system can make it easier for consumers to purchase products (Putra et al., 2019).

Based on the above problems, this research formulates the problem, namely: how to apply the a priori algorithm to determine the purchase pattern of clothing and accessories by consumers and how to apply the a priori algorithm to find out which products are predominantly chosen by consumers when buying products at PT. Surya Indah City.

*name of corresponding author



LITERATURE REVIEW

Data Mining

According to (Badrul, 2015) "Data mining is a technique which is a combination of methods of continuous data analysis with algorithms for processing large data.". According to (Anas, 2016), "Data mining algorithms mostly come from the form of developing algorithms in various fields of machine learning, statistics, artificial intelligence and artificial neural networks. Because it is not designed to handle data in very large sizes, while data mining in question is useful for handling data of such size, then one direction of research in the field of data mining is to develop these algorithms so that they can handle very large data sizes". According to (Handrianto & Farhan, 2019) "Data Mining is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and related knowledge from various large databases".

According to (Irfiani, 2019), "Data mining is a method by extracting large amounts of information, in this way helping companies focus on important information in data warehouses".

According to (Gunadi & Sensuse, 2012), "Data mining is the process of analyzing data to find a pattern from the data set". Data mining is able to analyze and process a set of data into information that can determine patterns that can support a decision. There are several algorithms in data mining that we can use to predict a result from processing a set of data. One of them is to use a priori algorithm. There are several algorithms in data mining that can be used to predict a result from processing a set of data, one of which is by using a priori algorithm.

Definition of Association Rule

According to (Panjaitan et al., 2020) states that, "analysis of association rules in data mining is a data mining technique for finding association rules between a combination of items". The association rule analysis is known as one of the data mining techniques that is the basis of various other data mining techniques. The analysis phase of association rules will produce an efficient algorithm using high frequency patterns. This association rule aims to look for patterns that often occur among many transactions, where each transaction consists of several items.

According to Kusriani in (Listriani et al., 2016) stated that "The basic methodology of association analysis consists of two, namely: analysis of high frequency patterns and the formation of associative rules".

a. High Frequency Pattern Analysis

This stage looks for combinations of items that meet the minimum requirements of the support value in the database. The value of an item's support is obtained by the following formula 1.

$$\text{Support (A)} = \frac{\text{Total Transaction Containing A}}{\text{Total Transaction}} * 100 \% \quad (1)$$

While the support values of the 2 items are obtained from the following formula 2.

$$\text{Support(A, B)} = \frac{\sum \text{Transaction Containing A and B}}{\sum \text{Transaction}} * 100 \% \quad (2)$$

Establishment of Associative Rules

The formation of the association rules is done after all high frequency patterns have been found, then the association rules are found to meet the minimum requirements for confidence by calculating the associative rule confidence "if A then B". The confidence value of the rule "if A then B" is obtained from the following formula3:

$$\text{confidence(A, B)} = \frac{\sum \text{TOTAL TRANSACTION CONTAINING A and B}}{\sum \text{TOTAL TRANSACTION A}} * 100 \% \quad (3)$$

Apriori Algorithm

A priori algorithm, including the type of association rules in data mining. Apart from a priori, which is included in this group are generalized rule induction methods and hash-based algorithms. A rule that states the association between attributes is often called an affinity analysis or market basket analysis. Association analysis or association rule mining is a data mining technique to find association rules between a combination of items (Wulandari, 2017).

According to (Rusdiansyah et al., 2020), Apriori algorithms include types of association rules in Data Mining. Rules that state the association between several attributes are often called affinity analysis or market basket analysis. Association analysis or association rule mining is a Data Mining technique to find the rules for a combination of items. According to (Badrul, 2016), "A priori algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 to determine frequent itemsets as a rule of boolean association. A priori algorithm including the type of association rules in data mining".

Broadly speaking, the work of a priori algorithms is:

- Formation of itemset candidates, k-itemset candidates are formed from a combination (k-1) -itemset obtained from the previous iteration. One feature of the Apriori algorithm is the pruning of k-itemset candidates whose subsets containing k-1 items are not included in the high frequency pattern with k-1 length.
- Calculation of support for each k-itemset candidate. Support from each k-itemset candidate is obtained by scanning the database to count the number of transactions containing all items in the k-itemset candidate. This is also a feature of the a priori algorithm which requires the calculation by scanning the entire database of the longest k-itemset.
- Set high frequency pattern. High frequency patterns containing k items or k-itemset are determined from candidate k-itemset whose support is greater than the minimum support. If no new high frequency pattern is obtained, the whole process is stopped. If not, then k plus one and return to part 1.

Tanagra

According to (Badrul, 2016), "Tanagra is a free and useful data mining software for academic purposes and can be taken from several data mining methods in the form of data exploration analysis, statistical learning, machine learning and database areas". Tanagra software is said to be free, because the software is open source where every researcher can access to the source code, and add his own algorithm, as long as the researcher agrees and complies with the software distribution license.

METHOD

Research Stages

The research methodology is a systematic stage to help well directed research. The following is the research methodology used in the research conducted by the author.

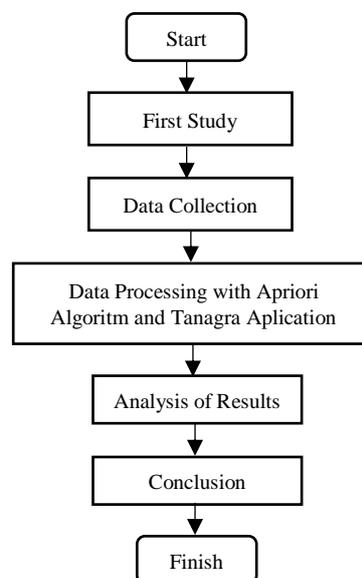


Figure 1. Research Stages

Data Collection Methods, Population and Research Samples

Method of Collecting Data

Research data collection was carried out in the following manner:

- Observation.** Pay attention to information related to sales transaction data at PT. Surya Indah City. The results of direct observation are then recorded so that it can be seen the application that can be done by the method that will be used.
- Interview.** Conducting direct interviews with the sales department of PT. Surya Indah City, who knows the product sales transaction data that occurred in the company for a year.
- Literature review.** Studying scientific journal literature and reference books to support research.

Population and Sample Studies

"Population is an area of generalization in the form of a subject or object under study and has certain qualities and characteristics that are determined by the researcher to be studied, then conclusions are drawn". (Sugiyono, 2015). Meanwhile, "The sample is partially or acts as a representative of the population so that the results of the research that were successfully obtained from the sample can be generalized to the population". (Sugiyono, 2015).

Population and Sample

The population which is the object of this research are some of the products sold at PT. Surya Indah City, where the total population is 9 products.

Table 1
Population

No	Product for Sale	Amount
1	9	9
Total		9

In this study, the samples that will be used to determine which products will be sold simultaneously are 9 products. This study uses nonprobability sampling, there are several types of samples in this study, the sample used for products sold simultaneously is saturated sample.

Table 2
Sample

No	Product for Sale
1	Clothes
2	Pants
3	Jackets
4	Bags
5	Hats
6	Skullcaps
7	Sandals
8	Shoes
9	Socks

Data Analysis Method

High Frequency Pattern Analysis

This stage looks for a combination of items that meet the minimum requirements of the support value in the database. The support value of an item is obtained by the following formula 1 :

$$Support(A) = \frac{\sum \text{Transaction containing A}}{\sum \text{Transaksi}} * 100\% \quad (4)$$

The support value of 2 items is obtained from the following formula 2 :

$$Support(A,B) = P(A \cap B)$$

$$Support(A,B) = \frac{\sum \text{Transaction containing A,B}}{\sum \text{Transaction}} * 100\% \quad (5)$$

Meanwhile, the support value of 3 items is obtained from the following formula 3 :

$$Support(A,B, \text{ and } C) = \frac{\sum \text{Transaction containing A,B and C}}{\sum \text{Transaction}} * 100\% \quad (6)$$

The formation of the association rule after all high frequency patterns are found, then we look for an associative rule that meets the minimum requirements for confidence by calculating the associative rule A - B The confidence value of rules A - B is obtained from the following formula :

$$Confidence = P(A|B) = \frac{\sum \text{Transaction containing A and B}}{\sum \text{Transaction A}} * 100\% \quad (7)$$

RESULT

Analysis Using the Apriori Algorithm Method

Apriori algorithms are commonly used to find association rules that meet the minimum requirements of support and confidence, which is the strength of the relationship between items in the association rule.

Establishment of 1 itemset

The process of completing the formation of 1 itemset or C_1 with a support amount of 30% with the formula :

$$= \frac{\Sigma \text{ Transaction containing A}}{\Sigma \text{ Transaction}} * 100\% \quad (8)$$

Table 3
1 Itemset

No	Itemset	Support
1	Clothes	66,7%
2	Shoes	58,3%
3	Socks	58,3%
4	Jackets	58,3%
5	Pants	58,3%

Establishment of 2 itemset

The process of completing the formation of 2 itemset or C_2 with a minimum amount of support of 30% with the formula :

$$\text{Support } (A,B) = P(A \cap B)$$

$$\text{Support } (A,B) = \frac{\Sigma \text{ Transaction containing A,B}}{\Sigma \text{ Transaction}} * 100\% \quad (9)$$

Table 4
2 Itemset

No	Itemset	Support
1	Clothes, Shoes	33,33%
2	Clothes, Pants	50%
3	Shoes, Socks	33,33%
4	Shoes, Jackets	33,33%
5	Socks, Jackets	33,33%

Establishment of Association Rules

After all high frequency patterns are found, then calculate the association rule with at least 60% confidence with the formula:

$$\text{Confidence} = P(A | B) = \frac{\Sigma \text{ Transaction containing A and B}}{\Sigma \text{ Transaction A}} * 100\% \quad (10)$$

Table 5
Candidates for Association Rules

Rule	Confidence
If you buy Clothes, you will buy Shoes	4/8 50%
If you buy Shoes, you will buy Clothes	4/7 57%
If you buy Clothes, you will buy Pants	6/8 75%
If you buy Pants, you will buy Clothes	6/7 85%
If you buy Shoes, you will buy Socks	4/7 57%
If you buy Socks, you will buy Shoes	4/7 57%
If you buy Shoes, you will buy Jackets	4/7 57%
If you buy Jackets, you will buy Shoes	4/7 57%
If you buy Socks, you will buy Jackets	4/7 57%
If you buy Jackets, you will buy Socks	4/7 57%

Final Association Rule

The final association rules are taken based on the candidate association rules that meet the requirements, namely with a minimum of 30% support and a minimum of 60% confidence.

Table 6
Final Association

Rule	Support	Confidence
If you buy Clothes, you will buy Pants	50%	75%
If you buy Pants, you will buy Clothes	50%	85%

Based on the table above, the most sold product is buying clothes, you will buy pants with 50% support and 75% confidence and if you buy pants, you will buy clothes with 50% support and 85% confidence. By knowing the many products that are sold, the company can use it for sales product layout.

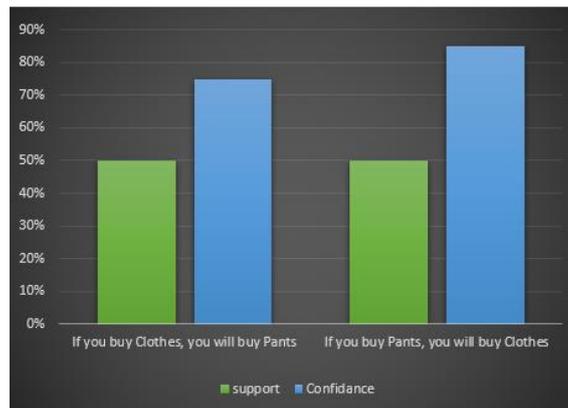


Figure 2. Final Association Rules Formation Result Diagram

DISCUSSIONS

Calculation with Tanagra

The algorithm for determining support can be seen below which consists of input, output and process. Following are the results of calculations using Tanagra,

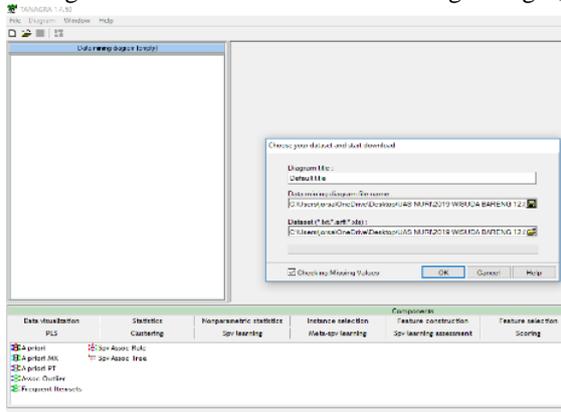


Figure 3. Input Process

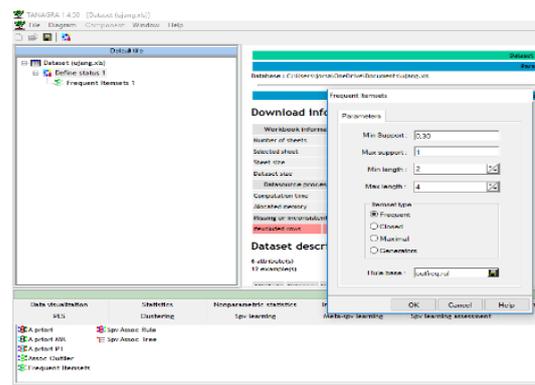


Figure 4. Fill in The Parameters Frequent Itemset

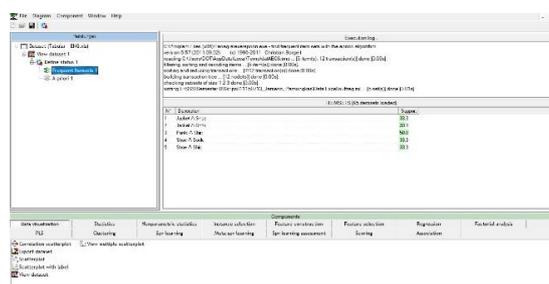


Figure 5. View Frequent Itemset

The confidence determination algorithm can be seen below which consists of input, output and process. If the confidence is less than 60% then the association rules are not included. Here are the results of calculations using Tanagra,

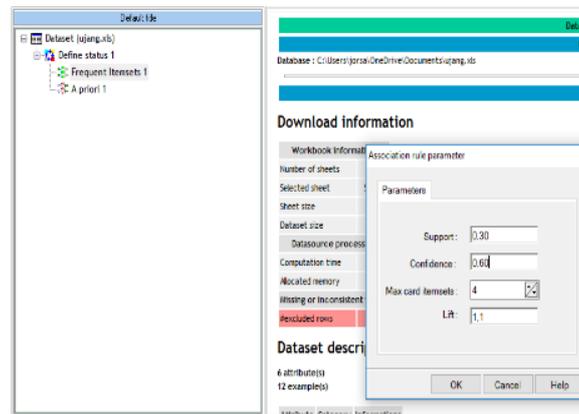


Figure 6. Parameters Apriori

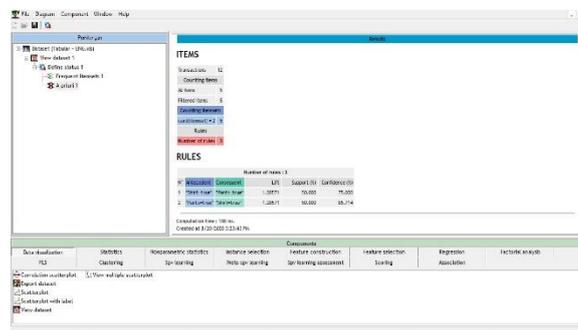


figure 3. Confidence Result

CONCLUSION

From the results of research that has been done to see the highest sales of products at PT. Surya Indah City by using the Apriori Algorithm method can be concluded that most of the sales of products sold at PT. Surya Indah City can be found using the Apriori Algorithm, by using 2 variables that meet support and minimum confidence, the most sold products are clothes and pants. From the final known association rule, if you buy clothes, you buy pants with 50% support and 75% confidence. If you buy pants, you are buying clothes with 50% support and 85% confidence. The priori algorithm can help develop a marketing strategy for marketing the marketed product by examining what are the advantages of the current most sold products

REFERENCES

Anas, A. (2016). Analisa Algoritma Apriori Untuk Mendapatkan Pola Peminjaman Buku Perpustakaan Smpn 3 Batanghari. *Jurnal Ilmiah Media SISFO*, 10(2), 628–641. <http://ejournal.stikom-db.ac.id/index.php/mediasisfo/article/view/233/220>

Badrul, M. (2015). Prediksi Hasil Pemilu Legislatif Dengan Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Pilar Nusa Mandiri*, XI(2), 152–160. <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/424/374>

Badrul, M. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. *None*, 12(2), 121–129. <https://media.neliti.com/media/publications/227549-algoritma-asosiasi-dengan-algoritma-apri-f4245cc8.pdf>

Fauziah, S., & Ratnawati. (2018). Penerapan Metode FIFO Pada Sistem Informasi Persediaan Barang. *Jurnal Teknik Komputer*, 4(1), 98–108.

Gunadi, G., & Sensuse, D. I. (2012). Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) : *Telematika*, 4(1), 118–132.

Handrianto, Y., & Farhan, M. (2019). C.45 Algorithm for Classification of Causes of Landslides. *Sinkron*, 4(1), 120. <https://doi.org/10.33395/sinkron.v4i1.10154>

Irfiani, E. (2019). Application of Apriori Algorithms to Determine Associations in Outdoor Sports Equipment

- Stores. *Sinkron*, 3(2), 218. <https://doi.org/10.33395/sinkron.v3i2.10089>
- Lestari, B. S. (2014). Fashion sebagai Komunikasi Identitas Sosial di Kalangan Mahasiswa. *Ragam Jurnal Pengembangan Humaniora*, 14(3), 225–238. <https://jurnal.polines.ac.id/index.php/ragam/article/view/514/>
- Listriani, D., Setyaningrum, A. H., & Eka, F. (2016). Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro). *Jurnal Teknik Informatika*, 9(2), 120–127. <https://doi.org/10.15408/jti.v9i2.5602>
- Panjaitan, L. F., Handrianto, Y., & Nurhadi, A. (2020). Apriori Algorithm On Car Rental Analysis With The Most Popular Brands. *Sinkron*, 4(2), 47. <https://doi.org/10.33395/sinkron.v4i2.10506>
- Putra, J. L., Raharjo, M., Sandi, T. A. A., & Prasetyo, R. (2019). Implementasi Algoritma Apriori Terhadap Data Penjualan. *Jurnal Pilar Nusa Mandiri*, 15(1), 85–90. <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/113/95>
- Rusdiansyah, Suharyanti, N., Triningsih, & Murniyati. (2020). Application Of Pizza Sales Data Mining Using Apriori Method. *Sinkron*, 4(2), 1. <https://doi.org/10.33395/sinkron.v4i2.10500>
- Santoso, H., Hariyadi, I. P., & Prayitno. (2016). Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori. *Teknik Informatika*, 1, 19–24. <http://ejournal.stikom-db.ac.id/index.php/mediasisfo/article/view/233/220>
- Sugiyono. (2015). Metode penelitian pendekatan kuantitatif, kualitatif dan R&D No Title. In *Metode penelitian pendekatan kuantitatif, kualitatif dan R&D* (pp. 80–84). Alfabeta.
- Wulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer* (1st ed.). Gava Media.