SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.10089
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of Apriori Algorithms to Determine Associations in Outdoor Sports Equipment Stores

Eni Irfiani
Universitas Bina Sarana Informatika
Jakarta, Indonesia
eni.enf@bsi.ac.id

***Abstract*—** A good sales strategy has an effect on increasing the number of sales of goods. Problems that often occur in outdoor sports equipment stores are the difficulty in determining sales strategies because there is not much interest in outdoor sports in the community. In addition, the amount of inventory in the store is excessive, which affects the sales cycle of goods. One way to help determine strategy is to use the apriori algorithm. In this method can determine consumer shopping behavior patterns. Apriori algorithms are part of the data mining analysis association. This algorithm is used to determine association rules. In the study, a combination of sports equipment purchased by consumers will be determined. Determination of the combination starts from 1 itemset to 3 itemsets, the combination of rule association produces different sales transaction patterns. The results of this research in the form of a combination of consumer shopping behavior patterns that will be used as recommendations for shop. Based on the results of the calculation of the support value of shoes and clothes has the largest value of 40.7% and a confidence value of 91.7%. The results of the rule association pattern for outdoor sports equipment stores can form a good marketing strategy by adding variants and brands to products that are of interest to customers and reducing the amount of inventory that is less attractive to customers.

Keywords—Apriori Algorithm; Association; Data Mining; Outdoor Sports Equipment

I. INTRODUCTION

The company is currently focusing more on old customers and selling more products than looking for potential new customers. The cost of getting new customers is more expensive than keeping old customers. Using the data mining method is able to accelerate the decision-making process and enable companies to manage information contained in transaction data (Tama, 2010).

Research on the reasons consumers buy products is very interesting to learn. To learn psychological thinking patterns can be done by analyzing consumer purchasing behavior. This research helps companies improve marketing strategies and reduce problems such as: how consumer spending patterns are influenced by the environment, the reason consumers buy different products, customer interest in products, how management increases marketing and marketing ideas that effectively reach consumers. The association rule method in data mining is used to extract data for very large industries (Raorane & Kulkarni, 2011).

Business competition in a free market requires businesses to find solutions and think of strategies for business continuity. When the process of making a sales strategy decision requires a lot of information resources to analyze the behavior of consumer spending patterns. The analysis is carried out by the company to find items that are most often purchased simultaneously from sales data. The association pattern search using the apriori algorithm produces an association rule to form a combination of items (Listriani, Setyaningrum, & A, 2016).

Mining algorithm of association rules is one of the most popular methods of emerge from large databases and get association rules so that new knowledge is discovered. Based on this algorithm, this study shows that this paper is the most efficient way to find frequent items, which is an increase in reducing the time spent scanning only a few

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.10089
e-ISSN : 2541-2019
p-ISSN : 2541-044X

transactions. The main components in data mining system architecture are: data warehouse, database, knowledge base, a server that is relevant from repositories, data mining engines. From the combination of patterns in the data mining module to get interesting patterns. Finally, users can communicate with data on data mining systems (Al-Maolegi & Arkok, 2014).

Data mining is a method by extracting large amounts of information, in this way helping companies focus on important information in data warehouses. The system proposed to find the most common item combination is the apriori algorithm. Research results are analyzed to make the right decision for the company (Shah, Solanki, Tambe, & Dhangar, 2015).

Data mining plays an important role in finding new trends in supermarket analysis so that it is beneficial for all parties related to this field. Extract data in data mining is done automatically and semi-automatically. Data mining consists of machine learning, database management; and artificial intelligence to extract new patterns for large data sets. Through data mining in supermarket applications, supermarket management data can be converted into knowledge management (Kaur & Kaur, 2017).

Customers are the main supporting factors of sales to require data analysis to increase the number of customers. The number of sales transactions is very important for the company's progress. A good strategy and promotion can increase the number of sales. Through data mining techniques, it is expected to find a pattern of taking goods by customers based on the most sales transaction data (Alfiah, Pandhito, Sunarni, Muharam, & Matusin, 2018).

When a store has more inventory than a consumer's demand, it is likely that the store will suffer a loss. In an effort to help overcome this problem a method in data mining is needed to resolve decision making. By using the apriori algorithm can produce rule association related to sales (Sholik & Salam, 2018).

Often research is done to help solve problems by digging information from customers. With methods in data mining, the apriori algorithm can look for patterns of purchase behavior carried out by customers and find out what items are purchased by customers simultaneously. Rule association is part of data mining to find associative rules for combinations of item sets (Septiana & Dharmayanti, 2016).

The apriori algorithm is a type of association in data mining that aims to find frequent item sets. The Apriori algorithm analysis is defined as a process to find all a priori rules that have fulfilled the minimum requirements for support values and minimum requirements for confidence values (Pane, 2013).

The problem that occurs in stores that specifically sell outdoor sports equipment is the lack of public interest in buying outdoor sports products compared to other sports products. In addition, there is often a large amount of excess inventory on certain products. By using the apriori algorithm, consumer spending patterns will be formed in the form of products that are often purchased by consumers.

II. LITERATURE REVIEW

2.1. Data Mining

Data mining is one of the most dynamic research techniques currently using database technology. To extract very large valuable data, an efficient way of exploring is needed. The process of extracting information that was previously useless and unknown became a large beneficial pattern (Mangla, Sarda, & Nadu, 2013).

Data mining is related to extracting large amounts of data. Some people call data mining with the term Knowledge Data Discovery (KDD). Data mining involves the integration of several fields such as databases, data warehouses, statistics, information seeking and so on (Han & Kamber, 2006).

The Cross-Industry Process for Data Mining Standard (CRISP-DM) is used to analyze data that is freely available and used to solve common problems in the fields of business and research. CRISP-DM has six phases, namely (Larose, 2005):

1. Business understanding phase
2. Data understanding phase
3. Data preparation phase
4. Modeling phase
5. Evaluation phase
6. Deployment phase

2.2. Association

Data mining has become the main method to obtain knowledge from data that has large amounts of data. Methods in data mining that can process large data between other classification algorithms, decision trees, association rules and, others. Association rules are used to find interesting relationships between attribute values and solid data and low threshold support (T & Sundaram, 2012).

Association of Information Systems and finally Produce Knowledge that can be trusted (Rodpysh, Aghai, & Madji, 2012).

Association rule mining is a technique for finding associative rules from combinations of item sets. In association rules, there are two parameters used, namely support and confidence. Support is a supporting value containing a number of items in a database, which strengthens the relationship between items in the rule association. Methodology analysis

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.10089
e-ISSN : 2541-2019
p-ISSN : 2541-044X

associations are divided into two ways (Han & Kamber, 2006) yaitu:

1. find all frequent item sets, this stage search for items combinations that often appear. Support values are obtained by the following formula:

$$\text{Support}(A, B) = P(A \cup B) \quad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

Value of confidence is obtained by the formula as follows:

$$\text{Confidence}\ (A \rightarrow B) = P(B \mid A)$$

$$\text{Confidence}\ (A \rightarrow B) = \frac{support\ (A \cup B)}{support\ (A)} \quad \ldots\ldots\ldots\ldots \quad (2)$$

2. Generate strong association rules from the frequent itemsets. After that, the frequency pattern is searched for, which has a strong level of itemsets dependence by calculating the associative rules.

$$\text{Confidence}\ (A \rightarrow B) = \frac{support_{count}(A \cup B)}{support_{count}(A)} \quad \ldots\ldots \quad (3)$$

To determine the rule association that will be chosen must be sorted by support and confidence which has the highest value. Rules are taken as many as n rules that have the greatest results..

III. PROPOSED METHOD

The analysis was used to determine the association of outdoor sports equipment stores using CRISP-DM. Research Flow Based on CRISP-DM which has six life cycle phases:

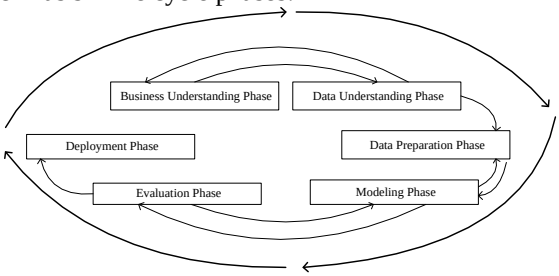

Figure 1. CRISP-DM flow

1. Business understanding phase. The purpose of sales at outdoor sports equipment stores is to provide equipment for consumers who want to do outdoor sports. lack of public interest in doing outdoor sports specifically hiking is still low compared to other types of sports. This affects the number of sales of sports equipment.
2. Data understanding phase.
   In this phase the data was collected through observation and interviews at the outdoor sports equipment stores in Bekasi. Sales data on the store used includes sales data online and offline.
3. Data preparation phase.
   In this phase data processing has been collected previously from sales transaction data. The data used as many as 200 sales transactions during the period January 2018 to August 2018, but only products with the highest sales value are selected based on frequency items.
4. Modeling phase.
   This study uses data mining techniques with the association rule method using the apriori algorithm. The results of this model are in the form of associative rules and consumer spending patterns.
5. Evaluation phase.
   In this phase, an evaluation model is used to make it effective and the quality of the model used and able to achieve the desired goals based on the value of support and confidence.
6. Deployment phase.
   The research results are represented in the form of information in a table of combination item sets to be taken into consideration in preparing a sales strategy.

How to solve problems using the apriori algorithm:

1. Determine items from sales transactions that have the highest frequency. From the transaction data for 8 months, the data that will be used is 54 sales data with the highest sales frequency.
   a. Determine the value of support for each of the products of clothes, pants, jackets, bags and, shoes. Determine support from a combination of 1 item set, 2 item sets and, 3 item sets. Item set combinations that have a support value less than the minimum support requirement of 30% will not be included in the rule association.
   b. Determine the confidence value of the item set combination that meets the minimum confidence requirement of 30%.
2. Determine the rule association of a combination of sets that meet the requirements and have the highest value.

IV. RESULT AND DISCUSSION

In the apriori algorithm calculation there are two ways:

1. Determine the frequent itemsets from transaction data. From the available transaction data, then the data is selected based on data that has the highest value and is purchased simultaneously.

Table 1. Transaction of products

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.10089
e-ISSN : 2541-2019
p-ISSN : 2541-044X

No_Trans	Date	Id. Cust	Item	Qty
1001035	05/01/2018	101	Pants	3
1001035	05/01/2018	101	Jackets	2
1002047	02/02/2018	102	Clothes	2
1002047	02/02/2018	102	Jacket	1
1002047	02/02/2018	102	Bags	3
1003056	13/03/2018	103	Clothes	2
1003056	13/03/2018	103	Pants	1
1003056	13/03/2018	103	Shoes	1
1005071	04/05/2018	104	Clothes	2
1005071	04/05/2018	104	Pants	1
1005071	04/05/2018	104	Jackets	1
1005071	04/05/2018	104	Shoes	3

Transaction data contained in table 1 contains sample sales transaction data based on daily transactions. After the transaction data is collected, the itemsets are calculated based on the value of support, so that the results of the formation of items can be obtained:

Table 2. Frequent 1-itemsets

No	Itemsets	Quantity	Support
1	Clothes	34	63%
2	Pants	29	54%
3	Jackets	34	63%
4	Bags	20	37%
5	Shoes	24	44%

After determining the value of 1 itemsets, then calculating the support value for the combination of 2 itemsetss, the results obtained:

Table 3. Frequent 2-itemsets

No	Itemsets	Quantity	Support
1	Pants, Jackets	18	33,3%
2	Pants, Bags	8	14,8%
3	Pants, Clothes	15	27,8%
4	Jackets, Bags	11	20,4%
5	Jackets, Clothes	24	44,4%
6	Clothes, Shoes	22	40,7%
7	Jackets, Shoes	13	24,1%

Based on table 3, there are two combinations that have more value than the minimum requirement of support values, namely (jackets, clothing) of 44.4%, (clothing, shoes) of 40.7% and, (pants, jackets) of 33.3% From the calculation, there is a minimum value of support of 30% if the combined value of items will be removed. Then the value of the 3 itemsetss combination is calculated, the results are:

Table 4. Frequent 3-itemsets

No	Itemsets	Quantity	Support
1	Pants, Jackets, Bags	3	5,6%
2	Pants, Jackets, Clothes	9	16,7%
3	Pants, Clothes, Shoes	10	18,5%
4	Pants, Bags, Clothes	0	0,0%
5	Jackets, Bag, Clothes	6	11,1%
6	Jackets, Shoes, Clothes	13	24,1%
7	Bag, Clothes, Shoes	0	0,0%

Based on table 4 of the combination of 3 itemsetss there is no value of support that exceeds the minimum support of 30%, then only a combination of 2 items sets meets the formation of rule associations. Determine the value of confidence from the combination of 2 itemsetss that have met the minimum confidence requirements. The result of the confidence value of the set combination formed as follows:

Table 5. Rule Association

Rule	Support	Confidence
Clothes, Shoes	40,7%	64,7%
Shoes, Clothes	40,7%	91,7%
Pants, Jackets	31,5%	59%
Jacket, Pants	31,5%	50%
Jackets, Clothes	38,9%	62%
Clothes, Jackets	38,9%	62%

2. Determine the rule association of a combination of sets that meet the requirements and have the highest value. After all high-frequency patterns are formed from a combination of 2 itemsetss in table 5, the next step is to find a rule association based on the highest value of support and confidence.

Table 6. Final Rule Association

Itemsets	Support	Confidence
Shoes, Clothes	40,7%	91,7%
Pants, Jackets	31,5%	59%
Jackets, Clothes	38,9%	62%
Clothes, Jackets	38,9%	62%

Based on the rule association in table 6, it is known that most customers buy shoes along with buying clothes, which have the highest value of support and confidence. While customers who buy jackets will buy clothes, as well as customers who buy clothes will buy a jacket too.

Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.10089
e-ISSN : 2541-2019
p-ISSN : 2541-044X

V. CONCLUSION AND SUGGESTION

From the results of the research, the application of a priori algorithms is concluded that the combination of 2 items sets produced 4 rule associations which form an association pattern. Where customers buy Shoes and Clothes have a support value of 40.7% while the confidence value is 91.7%. The results of the rule association pattern can be used in outdoor sports equipment stores in developing a good marketing strategy and can reduce the amount of inventory that is less attractive to customers.

Suggestion for outdoor sports equipment stores from the results of rule association is that if customers buy shoes, they will buy Jackets as well, shop owners can develop strategies such as preparing sales of shoes and jackets in the form of one package at a lower price. Store owners can prepare more variations and models of shoes and outdoor sports clothes of various types and brands compared to other products sold.

VI. REFERENCES

Al-Maolegi, M., & Arkok, B. (2014). An Improved Apriori Algorithm For Association Rules. *International Journal on Natural Language Computing (IJNLC)*, *3*(1), 21–29. https://doi.org/10.5121/ijaia.2011.2304

Alfiah, F., Pandhito, B. W., Sunarni, A. T., Muharam, D., & Matusin, P. R. (2018). Data Mining Systems to Determine Sales Trends and Quantity Forecast Using Association Rule and CRISP-DM Method. *International Journal of Engineering and Techniques*, *4*(1), 186–192. Retrieved from http://www.ijetjournal.org

Han, J., & Kamber. (2006). *Data Mining Concept and Techniques*. India: New Age International Limited.

Kaur, D., & Kaur, J. (2017). Data Mining in Supermarket: A Survey. *International Journal of Computational Intelligence Research*, *13*(8), 1945–1951. Retrieved from http://www.ripublication.com

Larose, D. T. (2005). *Discovering Knowledge in Data, an Introduction to Data Mining*. Canada: John Wiley & Sons, Inc.

Listriani, D., Setyaningrum, A. H., & A, F. E. M. (2016). Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro). *Jurnal Teknik Informatika Vol 9 No. 2, Universitas Islam Negeri Jakarta*, *9*(2), 120–127.

Mangla, V., Sarda, C., & Nadu, T. (2013). Improving the efficiency of Apriori Algorithm in Data Mining. *International Journal Of Science, Engineering And Technology*, *3*(3), 393–396.

Pane, D. K. (2013). Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus : Kreditplus). *Pelita Informatika Budi Darma*, *IV*(3), 25–29. https://doi.org/2301-9425

Raorane, A., & Kulkarni, R. V. (2011). Data Mining Techniques: A Source for Consumer Behavior Analysis. *International Journal of Database Management Systems*, *3*(3), 45–56. https://doi.org/10.5121/ijdms.2011.3304

Rodpysh, K. V., Aghai, A., & Madji, M. (2012). Applying Data Mining in Customer Relationship Management. *International Journal of Information Technology, Control and Automation*, *2*(3), 15–25. https://doi.org/10.5121/ijitca.2012.2302

Septiana, Y., & Dharmayanti, D. (2016). Penerapan Improved Apriori Pada Aplikasi Data Mining Di Perusahaan Kalvin Socks Production Jurnal Ilmiah Komputer dan Informatika (KOMPUTA). *Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA)*, *5*(1), 35–41.

Shah, N., Solanki, M., Tambe, A., & Dhangar, D. (2015). Sales Prediction Using Effective Mining Techniques. *International Journal of Computer Science and Information Technologies (IJCSIT)*, *6*(3), 2287–2289.

Sholik, M., & Salam, A. (2018). Implementasi Algoritma Apriori untuk Mencari Asosiasi Barang yang Dijual di E-commerce OrderMas. *Techno.Com*, *17*(2), 158–170.

T, S., & Sundaram, V. (2012). Association Models for Prediction With Apriori Concept. *International Journal of Advances in Engineering & Technology*, *5*(1), 2231–1963.

Tama, B. A. (2010). Penetapan Strategi Penjualan Menggunakan Association Rules dalam Konteks CRM. *Jurnal Generic*, *Vol. 5*(No.1), 35–38.