Online Store Product Recommendation System Uses Apriori Method

C S Fatoni¹, E Utami² and F W Wibowo³

¹,²,³ Magister of Informatics Engineering, AMIKOM Yogyakarta University, Indonesia.

E-mail: fatoni.work@gmail.com, emma@nrar.net, ferry.w@amikom.ac.id

Abstract. Online Store grows very fast. Online Store helps people to buy the desired product online. Heavy competition among of Online Store provider give a rise to technology development. Many Online Store system not only display the product but also need to be supported by proper products selection to attract the attention of website visitors. As a result, many website visitors are confused when they are going to buy products in Online Store. The number of product variety offered to a customer when he buys goods sometimes more than one product. The problem leads to an idea of developing a products recommendation system. Historical data from visitors and customers can be used to analyze the user needs and products preferences. Association rule using Apriori knowledge will be able to capture the user preference. By identify the user preferences, a valid product recommendation can be developed. This research will analyze the rules in a historical data of purchase from Online Store visitors to get recommendation of products to be displayed. According to experiment result, the association rule is capable to generate precise recommendations with confidence values 76.92%.

1. Introduction

The Internet becomes a very important medium for its users to share information and conduct business transactions by the company. Improving the service and business of the company becomes the main factor to develop online media as a means of conducting online transactions. The development of the field of online media business one of them is Online Store, it allows users to buy goods online. "Forget B2B (Business to Business) or B2C (Business to Customer), business is about P2P (Path to Profitability) - the path to profitability or profit" [1]. Information from around the World; eMarketer; about the Online Store share of total global retail sales from 2015 to 2021 is shown in Figure 1 [2].
The growing Online Store market growth is a major concern for business people to develop technology to make it easier for customers to buy the products they want. The growth of the digital economy the greater the business competition will also get heavier, consequently more and more Online Store not only display product products but also need to be supported by the selection of the right products to attract the attention of website visitors. The number of product variations offered to customers when purchasing goods is sometimes more than one product, consequently many customers are confused choosing the desired product, even there are customers who buy various products and per transaction.

The problem of confused customers when buying a product leads to the idea of developing a product recommendation system. Historical data from visitors and customers can be used to analyze user needs and product preferences. Association rules using Apriori knowledge will capture user preferences. By identifying user preferences, valid product recommendations can be developed. This research will analyze the rules in the historical data of Online Store visitor's purchase to get product recommendation to be displayed. Apriori method implements taking data with associative rules (Association rule) to determine the associative relationship of item combination [3]. Association rules are implemented through mechanisms of counting the support and trust of item relations. Association rules are interesting if the value of support is greater than the minimum support and also the value of the confidence is greater than the minimum trust. This a priori algorithm would be appropriate to apply when there are several link items to be analyzed.

Historical data about purchases from visitors and Online Store customers can be used to form item combination data, in testing data whether item combinations meet the minimum support and confidence parameters. When item combinations meet parameters, they can be used to determine valid product recommendations for customers who want to buy products.

2. Related Works

Section on the topic can be more than one according to the needs of the author. The content of the topic section according to the author's needs. For the section "topic", the author gives the title according to the needs of the author and the number of topics according to the needs of the author. If there is only one topic, then the section number follows to be: 1. Introduction, 2. Topic (title according to author's need), 3. Conclusion, 4. References.
The last remaining research (Ali Anaissi, and Madhu Goyal, 2015) raises the title of the SVM-Based Association Rules for Knowledge and Classification. The ultimate goal is to find an existing relationship between product sales that can help retailers identify new opportunities to sell their products to customers. This paper aims to find hidden patterns of knowledge in large data sets that can generate more understanding for data holders and identify new opportunities for important tasks including strategic planning and decision making. Researchers try to implement data mining tools that will allow analysts to interact with data and answer business questions such as promotional advertising. This research produces a list of association rules which are then grouped into different categories based on product results from the rules. SVM is used to test whether this group can capture variations based on different types of rules. Significant classification accuracy is achieved by SVM, which allows clients to target customers with their needs and propose a useful campaign [4].

Subsequent research (Ezhilarasan C, and Ramani S., 2017) research entitled Predicted performance using Modified Clustering Techniques with Fuzzy Association Rule Mining Approach to Retail classifies data based on components and related equations, using rules of mining fuzzy rules for implementation. Grouping is done on one object and set in one cluster, when overlapping and has multiple clusters for a new object using the fuzzy method. The size of the proximity is determined by the matrix. Analysis of this system is made to improve sales predictions better, and performance based on association rules using fuzzy model. In this study data were extracted from Online Store to predict sales by type of group [5].

Several studies related to information systems have been carried out by (Yusuke Ito, and Shohei Kato, 2016) research entitled An Apriori-Based Approach to Product Placement in Order Picking make and manage warehouse goods as easy as possible, the purpose is to shorten the time in the collection of goods in the warehouse, in this study written that the method applied is very successful and very effective, but this research is only done on small-scale warehouses, not on a large scale [6].

3. Methodology

The research has several stages ranging from literature studies to experimental methods and through the process of testing data mining. The system design method used is Reuse-Based has 6 stages in the system design process, among others, the collection of system specification requirements, component requirement analysis, system specification modification, combining the system design with Reuse-Based, development of merger system, and system validation process. The concept approach allows for the retrieval of reusable components and depends on the size of components that can be reused and integrated with the concept of components in the software, described in Figure 2 [7].

![Figure 2. Reuse Based Designing Method.](image)

3.1. System Requirement Collection

The process of collecting system requirements in this study covers the need for Online Store related information, data needs from problem identification, problem solving, database retrieval, data processing planning taken from database in realtime so that it will be processed into data mining.
3.2. Component Requirement Analysis
The analysis will be done after gaining access to the Online Store database then performing data retrieval periodically, with date filter, month by month. Grapheap data is a key component in data mining, if the data has been obtained then the steps to determine the database to store all the data, and then create a data matrix for the calculation using the a priori method.

3.3. Modification of System Specifications
Modification of system specifications that can be done in data mining, after making the process of data retrieval will be processed in a different database with existing Online Store system, it will create its own system integrated with Online Store.

3.4. Component Requirement Analysis
The application of Reused-Based is more to initialization of system requirements at the beginning, then the data retrieval is always updated, as well as the implementation of the process of retrieving historical data purchase transactions on Online Store on the system for data mining. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and gather knowledge from large databases or data warehouses [8]. Based on the definition that has been submitted, the important things related to data mining are data mining is an automated process for existing data, data to be processed in the form of very large data, and the purpose of data mining is to obtain relationships or patterns that may provide useful indications.

The next process is the analysis using the priori method aims to find the combination of items with the frequency pattern of the results of transactions on Online Store. To process the a priori method described in figure 3.

![Figure 3. Reuse Based Designing Method.](image)

In Figure 3 the priori method will continue if the data condition is still possible for the next itemset. If not possible then soon be the result. The customer purchase transaction matrix will take the number of transactions from each product item per transaction and the amount of transaction data then used to determine the combination of product items.

Apriori method is one algorithm that searches for frequent itemset by using rule association technique. The Apriori method uses knowledge of the frequency of previously known attributes for processing more information. In the Apriori algorithm determines possible candidates with respect to minimum support and minimum trust. Support is the visitor value or the percentage of item combinations in the database.

The main processes undertaken in the Apriori method for obtaining frequent itemsets include: (1) Join (merge): This process is done by combining items with other items until no more combinations can be
established; and (2) Prune (pruning): The trimming process is the result of a merged item then trimmed by using the minimum support specified by the user. The principles of the Apriori Algorithm include: (1) Collect single items then look for the biggest item; (2) Get a candidate pair and then count the major pair of each item; (3) Find the candidate triplets of each item and so on; and (4) Every part of the frequent itemset should be frequent.

3.5. System Development and Integration
The system development process and its integration involve retrieving data for processing in data mining and then applying a priori method to determine the combination of item item items, which will then be processed by a system integrated with Online Store and displayed in a data format as per system requirements analysis for recommendations customer products.

3.6. System Validation Process
System validation refers to the system test of the data processed in this study compared to the data processed by using the system created, for validation of the desired results will be tested gradually by using more data to obtain recommendations from more results.

4. Result and Discussion

4.1. System Design Analysis
The process of system design analysis in accordance with the steps in the flow of Online Store product recommendation system, following the flow of system design analysis process described there is Figure 4.

![Figure 4. Design of Process Flow Analysis System](image)

The process of system design analysis begins with data collection related to customer purchase transactions in Online Store. The data obtained are then analyzed and presented in the historical data of Online Store visitor purchases, with associative rules for determining the associative relationship of a combination of product items purchased by the customer, the calculation method using the mechanism of counting the support and beliefs of a relationship item. Association rules are interesting if the value of support is greater than the minimum support and also the value of the confidence is greater than the minimum trust. Apriori algorithm is applied when there are several items of relationships that want to be analyzed. Testing data mining processing then from the results of product output recommendations appropriate for customers will be done gradually by looking at the next customer purchase history data in a certain period.

4.2. Data collection
The initial process of system design analysis is to collect customer purchase history data taken at the end of the year from June to July 2017.

Table 1. Pattern of Customer Purchase Transaction

| Trx | Item Transaction                          |
|-----|------------------------------------------|
| 1   | Shoes, Jacket, Pants, Shirts, Hijab      |
| 2   | Hands, Bags, T-shirts, Shirts, Pants     |
| 3   | Clothes, Shoes, Bags, Wallets, Pants     |
| 4   | Perfumes, Shirts, Bags, Glasses, Shoes   |
| 5   | Wallet, Bag, Jacket, Jas, Perfume        |
4.3. Transaction Data Matrix

Customer purchase transaction data will be adjusted in matrix format to get the pattern of goods formed, and used to know how many times the goods are purchased by the customer on each transaction, the following data is presented in table 2.

Table 2. Customer Purchase Transaction Matrix

| Trx | Item transactions |
|-----|-------------------|
|     | Shoes | Jacket | Pants | Wristwatch | Bag | T-shirts | Shirt | Perfume | Eyeglasses | Purse | Belt | Jas | Sweater | Hijab | Dress |
| 1   | 1     | 1      | 0     | 0         | 0   | 0        | 1     | 0        | 0          | 0     | 0    | 1   | 0       | 0     | 0     |
| 2   | 0     | 0      | 1     | 1         | 1   | 1        | 1     | 0        | 0          | 0     | 0    | 0   | 0       | 0     | 0     |
| 3   | 1     | 0      | 0     | 0         | 1   | 0        | 1     | 0        | 0          | 1     | 0    | 0   | 0       | 0     | 0     |
| 4   | 1     | 0      | 0     | 0         | 1   | 0        | 1     | 1        | 1          | 0     | 0    | 0   | 0       | 0     | 0     |
| 5   | 0     | 1      | 0     | 0         | 1   | 0        | 1     | 0        | 1          | 1     | 0    | 0   | 0       | 0     | 0     |
| 6   | 0     | 0      | 0     | 0         | 0   | 0        | 1     | 1        | 1          | 0     | 0    | 1   | 0       | 0     | 0     |
| 7   | 0     | 0      | 0     | 0         | 0   | 0        | 0     | 0        | 0          | 0     | 0    | 0   | 0       | 1     | 1     |
| 8   | 1     | 0      | 1     | 0         | 0   | 0        | 1     | 1        | 0          | 0     | 0    | 1   | 0       | 0     | 0     |
| 9   | 1     | 0      | 1     | 0         | 0   | 0        | 0     | 1        | 0          | 0     | 0    | 1   | 0       | 0     | 0     |
| 10  | 1     | 0      | 1     | 0         | 1   | 1        | 0     | 0        | 0          | 1     | 0    | 0   | 0       | 0     | 0     |
| 11  | 0     | 1      | 1     | 0         | 0   | 1        | 0     | 0        | 0          | 0     | 0    | 1   | 1       | 0     | 1     |
| 12  | 0     | 0      | 1      | 1         | 0   | 1        | 0     | 0        | 0          | 0     | 0    | 1   | 1       | 1     | 1     |
| 13  | 1     | 0      | 1      | 0         | 0   | 0        | 1     | 0        | 0          | 1     | 0    | 1   | 1       | 0     | 0     |
| 14  | 1     | 0      | 0      | 0         | 1   | 1        | 1     | 0        | 0          | 0     | 0    | 0   | 1       | 0     | 0     |
| 15  | 1     | 0      | 0      | 0         | 0   | 0        | 0     | 0        | 1          | 1     | 1    | 0   | 0       | 0     | 0     |
| 16  | 1     | 0      | 1      | 0         | 0   | 0        | 0     | 0        | 1          | 0     | 1    | 0   | 0       | 0     | 0     |
| 17  | 1     | 0      | 1      | 1         | 0   | 0        | 0     | 0        | 0          | 0     | 0    | 0   | 1       | 0     | 0     |
| 18  | 1     | 0      | 0      | 1         | 0   | 0        | 0     | 0        | 0          | 0     | 0    | 0   | 1       | 1     | 1     |
| 19  | 1     | 0      | 1      | 1         | 1   | 0        | 0     | 0        | 0          | 0     | 0    | 0   | 0       | 0     | 1     |
| 20  | 0     | 0      | 1      | 0         | 0   | 0        | 1     | 1        | 1          | 1     | 0    | 0   | 0       | 0     | 0     |
4.4. Analysis of Itemset Creation

The customer purchase transaction matrix will take the number of transactions from each product item per transaction and the amount of transaction data then used to determine the itemset combination.

4.4.1. Itemset Combination Analysis. The combination of 1 itemset is processing based on the data provided in table 2, the process of forming $K_1$ or called a combination of 1 itemset with the minimum amount of support = 40%, by the formula in equation (1) [9].

$$\text{Support (item)} = \frac{\text{Number of transactions containing Item}}{\text{Total transactions}}$$

(1)

The squatter process in equation (1) obtains the data shown in table 3, for the support value of each product item.

| Item       | Quantity | Support |
|------------|----------|---------|
| Shoes      | 13       | 65%     |
| Jacket     | 3        | 15%     |
| Pants      | 13       | 65%     |
| Wristwatch | 6        | 30%     |
| Bag        | 10       | 50%     |
| T-shirts   | 6        | 30%     |
| Shirt      | 12       | 60%     |
| Perfume    | 5        | 25%     |
| Eyeglasses | 5        | 25%     |
| Purse      | 7        | 35%     |
| Belt       | 3        | 15%     |
| Jas        | 3        | 15%     |
| Sweater    | 3        | 15%     |
| Hijab      | 9        | 45%     |
| Dress      | 4        | 20%     |

The establishment of itemsets in table 3 with a minimum of 40% support can be found that meets the minimum standards of support on shoe product items, pants, watches, bags, shirts, wallets. Then from the result of combination formation 1 item will be done combination 2 itemset as in table 4.

4.4.2. Itemset Combination Analysis. The Combination of 2 Items is processing based on data provided in table 3 items taken above the support value of each product item, the process of forming $K_2$ or called a combination of 2 itemsets with minimum amount of support = 40%, by the formula in equation (2) [9].

$$\text{Support (P \cap Q)} = \frac{\sum \text{transactions contain P \& Q}}{\sum \text{transactions}}$$

(2)

The squatter process in equation (1) obtains the data shown in table 3, for the support value of each product item.

| Item          | Quantity | Support |
|---------------|----------|---------|
| Shoes, Pants  | 9        | 45%     |
| Shoes, Bags   | 5        | 25%     |
| Shoes, Shirts | 8        | 40%     |
| Shoes, Hijab  | 6        | 30%     |
| Pants, Bags   | 5        | 25%     |
| Item               | Quantity | Support |
|-------------------|----------|---------|
| Pants, Shirts     | 10       | 50%     |
| Pants, Hijab      | 6        | 30%     |
| Bags, Shirts      | 4        | 20%     |
| Bags, Hijab       | 4        | 20%     |
| Clothes, Hijab    | 4        | 20%     |

Combination 2 itemsets with minimum 40% support can be seen as a combination of 2 itemsets that meet the minimum standard of support that is shoes, pants with a support of 45% and pants, shirt with 50% support. From the result of the combination of 2 itemsets, the formation of 3 itemsets will be done as shown in Table 5.

### 4.4.3 Itemset Combination Analysis
Combination of 3 Items is processed based on data provided in Table 4 and items taken above the support value of each product item. The formation process K3 or called with a combination of 3 itemsets with minimum support = 40%, by the formula in equation (3) [10].

\[
\text{Support} (P, Q, R) = \frac{\sum \text{transactions contain } P, Q \text{ and } R}{\sum \text{transactions}}
\]  

(3)

**Table 5.** Support value from a combination of 3 product items

| Item                | Quantity | Support |
|---------------------|----------|---------|
| Shoes, Pants, Shirts| 6        | 30%     |

### 4.5. Association Rules
After all the high frequency patterns are found, then the association rules that meet the minimum requirements for confidence by calculating the trust of the associative rule P → Q [10]. The minimum Confidence = 70%. The confidence value of rule P → Q is obtained.

\[
\text{Confidence} = (Q|P) = \frac{\sum \text{transactions contain } P \& Q}{\sum \text{transactions contain } P} 
\]

(4)

**Table 6.** Support value from a combination of 3 product items

| Rules               | Quantity | Confidence |
|---------------------|----------|------------|
| If buying shoes, buy pants | 9/13     | 69.23%     |
| If buying a pair of pants, buy a T-shirt | 10/13    | 76.92%     |

Based on Table 6, the products most often purchased by customers are shoes, pants, T-shirts with knowledge of the products most often purchased by customers, then the company can develop strategies in determining the purchase of products to maintain product availability required by customers and also can adjust the location of the product based on the combination of product items formed.

### 4.6. System Implementation
The authors of this study took sample data from June to July 2017 to provide product recommendations in December 2018. Based on this transaction, the authors took the highest trust value as the basis for determining the right product recommendations for Online Store customers. Figure 5 shows the product brand and the type of product being sold.
Figure 5. Product Form and Product Data Brand

Figure 6. Product Transaction Form by Product Type

Figure 6 shows the transaction data of the type of product purchased from June to July 2017. From this product type data by applying the a priori method by applying a minimum support value of 40%, and a minimum value of 70% confidence generating confidence values from various itemset combinations. Based on the value of itemet combination with the highest confidence value, the system can provide product recommendations to be promoted or sold to Online Store customers in Figure 7.

Figure 7. Output Form Generate Product Recommendations
4.7. System Testing
Calculation with priori method in table 6 and from calculation with built system is determined minimum value of support: 40% and minimum trust value 70% then generated tendency of product purchased by customer is presented in table 7.

### Table 7. The calculation results

| Calculation Technique                  | Results                                      |
|----------------------------------------|----------------------------------------------|
| Calculation using Apriori method       | If buying a pair of pants then buy a t-shirt with a trust value of 76.92% |
| Calculations using the built system    | If buying a pair of pants then buy a t-shirt with a trust value of 76.92% |

Testing using a priori method and using the built system yields the same value in accordance with the minimum value of support and minimum trust that has been determined. From these results then this system can help the pattern of product display recommendations based on the tendency of Online Store products purchased by customers consisting of 1 item of the product, then from the test results can also help management Online Store in providing product recommendations to customers and as desired.

5. Conclusion

Based on the discussion of research that the authors do, it can be concluded that by applying the a priori algorithm, the system provides product recommendations to Online Store customers based on the trust value of a combination of products purchased at a given time period. Application of Apriori Method in this research is to find the most combination of items based on transaction data and then form the association pattern of item combination. This research can be further developed using a weighted product method to get the weight of each product brand so that it can provide an alternative to the most recommended product brands.

Acknowledgments

Universitas AMIKOM Yogyakarta for the support by giving the facility which is needed for the study. This research also can be done by a good cooperation among all sector.

6. References

[1] Berman D K 2000 *Dot Coms: Can They Climb Back?* (Business Week online June 26th)
[2] eMarketer 2018 *Information from around the World: 2015 to 2017* ([https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide](https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide))
[3] Aprizal, Hasriani, and Wahyu N 2015 *Implementasi Data Mining Untuk Menentukan Posisi Barang Pada Rak Menggunakan Metode Apriori Pada PT. Midi Utama Indonesia* (Techno. COM) Vol 15 p 335-342.
[4] Anaissi A and Goyal M 2015 *SVM-Based Association Rules for Knowledge Discovery and Classification, 2015 2nd Asia-Pacific World Congress on Computer Science and Engineering* (APWC on CSE)
[5] Ezhilarasan C and Ramani S 2017 *Performance Prediction using Modified Clustering Techniques with Fuzzy Association Rule Mining Approach for Retail 2017 International Conference on Intelligent Computing and Control*
[6] Ito Y and Kato S 2016 An Apriori-Based Approach to Product Placement in Order Picking
International Conference on Agents

[7] Fatoni C S and Noviandha F D 2018 Case Based Reasoning Diagnosis Penyakit Difteri dengan
Algoritma K-Nearest Neighbor Citec Journal Vol 4 No 3 p 220-232

[8] Turban E 2005 Decision Support Systems and Intelligent Systems (Yogyakarta: Andi Offset)

[9] Yanto R and Khoiriah R 2016 Implementasi Data Mining dengan Metode Algoritma Apriori
dalam Menentukan Pola Pembelian Obat Citec Journal Vol 2 No 2 p 102-113

[10] Ummi K 2016 Analisa Data Mining Dalam Penjualan Sparepart Mobil Dengan Menggunakan
Metode Algoritma Apriori (Medan: CSRID) Vol 8 No 3 p 145-208