Consumer purchase patterns based on market basket analysis using apriori algorithms

A R Efrat¹, R Gernowo² and Farikhin³

¹Information System Department, Faculty of Postgraduate School, Diponegoro University, Semarang, Indonesia
²Physics Department, Faculty of Mathematics and Natural Sciences, Diponegoro University, Semarang, Indonesia
³Mathematics Department, Faculty of Mathematics and Natural Sciences, Diponegoro University, Semarang, Indonesia

Corresponding author: rezhaefrat22@outlook.com

Abstract. Analyzing customer purchasing patterns can help minimarket expand marketing strategies by gaining insight into which items are often bought together by customers. Also, transaction data is a source of information available at the convenience store and one thing that can be used for business decision making. In this paper, we aim to use the Apriori algorithm method to obtain consumer purchasing patterns to analyze consumer purchasing patterns. This system uses a priori algorithm calculation method where the input data is consumer transaction data. Transaction data will be sorted and calculated by providing a minimum support value and a minimum confidence value. In this study, the authors conclude that the results of the analysis of information systems in determining consumer purchasing patterns can be as information to determine sales and in the application of Apriori algorithms can provide information in the form of a combination of consumer purchase patterns based on consumer transaction data with a minimum support value above 10% and the minimum confidence value is above 65%.

1. Introduction

One kind of business with a high level of competition in the retail industry. In the retail business, success is influenced by the speed of response and the capability to understand consumer behavior. Also, a common problem in the retail business is to identify which items are most sold and which items tend not to be sold based on conformity with previous customer purchasing patterns [2]. Analyzing customer purchasing patterns can help minimarket increase strategies of marketing by getting insight into which items are consumers often buy [3]. Also, transaction data is an available source of information available at the minimarket and one thing that can be used for business decision making[4], [5]. Most of the sales transaction data are not reused, only stored as an archive and used to make sales reports [5].

The application of data mining techniques is indispensable in business objectives in retail companies such as Minimarket. Minimarkets can analyze customer shopping habits to discover associations and correlations between items in the shopping cart [6]. By implementing data mining techniques, it is expected to provide benefits for companies to continue to improve their business strategies by studying consumer behavior patterns [7]. Most companies, such as minimarket use data mining techniques in transaction databases to get the knowledge that helpful, which can support decision making to increase interactions between organizations and customers [8].
One of the most popular association rule mining is the Apriori algorithm. Apriori algorithm is used to study the association rules concept in basket market analysis. [9]. In the process, the Apriori algorithm can find high-frequency patterns that have a frequency above a certain threshold in a database. The iterative process will be performed until the algorithm cannot discover a new itemset and produces a robust association rule, with an "if-then" pattern based on minimum confidence and minimum support. [10]. In this paper, we aim to the analysis of consumer purchase patterns based on analysis of market basket using the Apriori algorithm and the data used are transaction data obtained in case studies.

2. Literature and method review

2.1. Analysis of market basket

Analysis of Market Basket aims to set what items are most often bought or items that consumers use most often [9]. Market Basket Analysis analyzes purchase habits from consumers and to discover relationships between different products that consumers place in basketball shopping [3]. Market basket analysis is a process that analyzes the habits of buyers to find relationships between various items in their shopping basket[2]. The finding results can be utilized by companies of retail such as a minimarket in developing marketing strategies to view goods that can be purchased simultaneously by consumers [10] and provide support for the scientific decision-making process for minimarkets through association relationships in data mining between items purchased [11].

2.2. Rules of association

Rules of the association are a manner to discover relationships between an item and other items. Rules of the association are usually implemented "if" and "then". To determine the rules of the Association, it is necessary to determine the support and confidence to limit whether the resulting rules are attractive or not [10]. Rules of the association can be applied to carry out marketing promotions, for example, by arranging product layouts on minimarket shelves precisely and by giving suggestions directly to customers of interesting items [12]. The analysis of association methodology is split into 2 phase, as follows [5]:

2.2.1. Analysis of the highest frequency pattern. This phase finds for combinations of items that fulfill the minimum needs support value of an item's. Support is containing items combination in percentage compared to the total number of transactions. In other words, support is the probability of a transaction containing items A and B divided by the whole transaction. So the calculation of support can be done with equations 1 and 2 as follows [13]:

\[
\text{Support} (A) = \frac{\sum (\text{Transaksi yang mengandung } A)}{\sum \text{Transaksi}} \tag{1}
\]

While the support values for the 2 items are obtained from the following formula:

\[
\text{Support} (A, B) = \frac{\sum (\text{Transaksi yang mengandung } A \text{ and } B)}{\sum \text{Transaksi}} \tag{2}
\]
2.2.2. Rules of association formation. After the highest frequency pattern is discovered, then find for rules of association that fulfill the minimum confidence. The higher the value of confidence, the greater the likelihood that goods will be purchased [3]. Confidence of association rules is that all transactions containing A and B are divided by all transactions that contain A. Confidence calculation can be done with the following equation 3.

\[
\text{Confidence} = P(B | A) = \frac{\sum \text{(Transactions containing A and B)}}{\sum \text{(The Total Processing A)}}
\]

2.3. The lift ratio

The Lift Ratio shows the power possessed by the rules of association result. [14]. Overall, the Lift Ratio summarizes the strength of associations between products [3]. The formula of the Lift Ratio can be seen in equation 4.

\[
\text{Lift Ratio} = \frac{\text{Confidence} \ (A \cap B)}{\text{Support} \ (A) \cdot \text{Support} \ (B)}
\]

2.4. Apriori algorithm

One of the rules of association in data mining is the Apriori algorithm. Examples of the rules of association implementation are conducting a purchase analysis at a grocery store to be able to find out how likely it is that customers buy items simultaneously [16]. Based on this knowledge, grocery store owners can arrange their own product placement on store shelves or design marketing campaigns utilizing a coupon of discount combinations for specific items [10]. An outline of the Apriori algorithm works[15]:

- Itemset candidates Formation, candidates of k-itemset are formed from a combination of items (k-1)-itemset are got from the previous iteration. Apriori algorithm characteristic is the pruning candidates of k-itemset whose subset contains k-1 items that not included in the pattern of high-frequency with the length of k-1.
- Support calculation for each candidate of k-itemset. For each candidate, Support is got by scan the database to count the transaction number containing all items are made as candidates. The algorithm involves calculations by the database scanning the with the items are the longest.
- Set pattern with high-frequency. Pattern with high-frequency containing k-items or item sets is specified from candidates who Support \( \geq \) the support of minimum.
- the process is stopped if no patterns of high-frequency are found again. Otherwise, then K+1 and back to part 1.

3. Result and discussion

3.1. Market basket analysis

The data used for this research, namely minimarket transaction data obtained from case studies is a history of purchases made by customers. However, in this paper, the transaction data used are only 20 transaction data that started from 1 June-2019 until 20 June-2019. After transaction data is presented as table 1, the transaction data is modeled by applying the Apriori Algorithm to look for the associated association rules, then look for support and confidence values.

| No | Date Transaction | Produk |
|----|------------------|--------|
| 1  | 01/June/2019     | Aice Sweet Corn 52gr, Fruit Tea Black Currant 500ml, Bear Brand 189ml, Dunhill Fine Cut Filter Black, Aqua Botol 1500ml |
| 2  | 02/June/2019     | Nescafe Coffe Cream 200ml, Prima Air Mineral 600ml |
| 3  | 03/June/2019     | Aqua Botol 1500ml, Aice Sweet Corn 52gr, Gg Signature Cok 12, Dunhill Fine Cut Filter Black |

Table 1. Consumer purchase transaction data
3.1.1. Apriori algorithm analyst. Transaction data in table 1, the association rule calculation is performed using the Apriori algorithm to produce rules that have support and confidence values. The results of calculations with formulas 1 and 2 can be seen in table 2. Based on the steps in the calculation process of the association rules [15], the following results are obtained:

- Calculation on K1 (candidate 1-internet) with support value > 10% and confidence value > 65%.

| No | Date Transaction | Produkt                                                                 |
|----|------------------|--------------------------------------------------------------------------|
| 4  | 04/June/2019     | Aqua Botol 1500 ml, Bear Brand 189ml, Cap Kaki Tiga Lychee 320ml          |
| 5  | 05/June/2019     | Aice Sweet Corn 52gr, Aice Strawberry Crispy 55gr, Prima Air Mineral 1500ml, Men's Biore FF Acne Def 40gr, Dunhill Fine Cut Filter Black, Sprite Botol 1 Liter, Teh Pucuk 350ml |
| 6  | 06/June/2019     | Dunhill Fine Cut Pth 20, Dunhill Fine Cut Filter Black, Gg Surya Pro Mild 16, You C 1000 Vit Lemon 140ml, Prima Air Mineral 600ml |
| 7  | 07/June/2019     | Aqua Botol 330ml, You C 1000 Vit Lemon 140ml, Aqua Botol 1500ml, Sampoerna A Mild 12 |
| 8  | 08/June/2019     | Aice Strawberry Crispy 55gr, Prima Air Mineral 600ml, Gg Signature Cok 12 |
| 9  | 09/June/2019     | Dunhill Ultra 12, Aqua Botol 600ml, Magnum Mild 234 Kretek 16 |
| 10 | 10/June/2019     | Gg Surya Pro Mild 16, Piattos Sapi Panggang 12gr, Aqua Botol 600ml, Kapal Api Special Merah 165g, Piattos Sapi Panggang 12gr |
| 11 | 11/June/2019     | Cimory Rasa Mangga 250ml, Chicken Eggs, Beef Sausage Gaga, Aqua Botol 1500ml |
| 12 | 12/June/2019     | Dunhill Fine Cut Putih 20, Aqua Botol 1500ml, Bear Brand 189ml, Aqua Botol 600ml |
| 13 | 13/June/2019     | You C 1000 Vit Lemon 140ml, Aqua Botol 600ml, Cap Kaki Tiga Lychee 320ml |
| 14 | 14/June/2019     | Magnum Mild 234 Kretek 16, Prima Air Mineral 600ml, Nescafe Coffe Cream 200ml |
| 15 | 15/June/2019     | Chicken Eggs Beef Sausage Gaga, Roma Biskuit Kelapa 450gr, Aqua Botol 600ml |
| 16 | 16/June/2019     | Aqua Botol 600ml, Cap Kaki Tiga Lychee 320ml, Roma Biskuit Kelapa 450gr |
| 17 | 17/June/2019     | Rinso Molto 770gr, Sunlight Ref Jeruk Nipis 230ml, Gg Signature Cok 12 |
| 18 | 18/June/2019     | Aice Strawberry Crispy 55gr, Cap Kaki Tiga Lychee 320ml, Crystalline 600ml, Aice Sweet Corn 52gr |
| 19 | 19/June/2019     | Ultra Milk Coklat 250ml, Roma Biskuit Kelapa 450gr, Teh Pucuk 350ml, Kapal Api Special Merah 165g |
| 20 | 20/June/2019     | Sprite Botol 1 Liter, Piattos Sapi Panggang 12gr, Cheetos Jagung Bakar 75g, Teh Pucuk 350ml, Crystaline 600ml |
| No | Item                        | Amount | Candidate (%) |
|----|-----------------------------|--------|---------------|
| 7  | Gg Signature Cok 12         | 3      | 3/20 * 100 = 15 |
| 8  | Cap Kaki Tiga Lychee 320ml  | 4      | 4/20 * 100 = 20 |
| 9  | Aice Strawberry Crispy 55gr | 3      | 3/20 * 100 = 15 |
| 10 | Sprite Botol 1 Liter        | 2      | 2/20 * 100 = 10 |
| 11 | Teh Pucuk 350ml             | 3      | 3/20 * 100 = 15 |
| 12 | Dunhill Fine Cut Pth 20     | 2      | 2/20 * 100 = 10 |
| 13 | Gg Surya Pro Mild 16        | 2      | 2/20 * 100 = 10 |
| 14 | You C 1000 Vit Lemon 140ml  | 3      | 3/20 * 100 = 15 |
| 15 | Aqua Botol 600ml            | 6      | 6/20 * 100 = 30 |
| 16 | Magnum Mild 234 Kretek 16   | 2      | 2/20 * 100 = 10 |
| 17 | Piattos Sapi Panggang 12gr  | 2      | 2/20 * 100 = 10 |
| 18 | Kapal Api Special Merah 165g| 2      | 2/20 * 100 = 10 |
| 19 | Chicken Eggs                | 2      | 2/20 * 100 = 10 |
| 20 | Beef Sausage Gaga           | 2      | 2/20 * 100 = 10 |
| 21 | Roma Biskuit Kelapa 450gr   | 3      | 3/20 * 100 = 15 |
| 22 | Crystaline 600ml            | 2      | 2/20 * 100 = 10 |

- Candidate 2-itemset can be seen in table 3.

| No | Item 1                          | Item 2                          | Candidate (%) |
|----|---------------------------------|---------------------------------|---------------|
| 1  | Aice Sweet Corn 52gr            | Dunhill Fine Cut Filter Black    | 3/20 * 100 = 15 |
| 2  | Aice Sweet Corn 52gr            | Aice Strawberry Crispy 55gr      | 2/20 * 100 = 10 |
| 3  | Bear Brand 189ml                | Aqua Botol 1500ml               | 2/20 * 100 = 10 |
| 4  | Nescafe Coffee Cream 200ml      | Prima Air Mineral 600ml          | 2/20 * 100 = 10 |
| 5  | Cap Kaki Tiga Lychee 320ml      | Aqua Botol 600ml                | 2/20 * 100 = 10 |
| 6  | Sprite Botol 1 Liter            | Teh Pucuk 350ml                 | 2/20 * 100 = 10 |
| 7  | Aqua Botol 600ml                | Roma Biskuit Kelapa 450gr       | 2/20 * 100 = 10 |
| 8  | Chicken Eggs                    | Beef Sausage Gaga               | 2/20 * 100 = 10 |

3.1.2. Final association rules with apriori algorithms. The results of the rules of the association are based on the minimum confidence and minimum support that has been fulfilled as follows:

| No | Itemset                                                                 | Support (%) | Confidence (%) | Lift Ratio (%) |
|----|-------------------------------------------------------------------------|-------------|----------------|----------------|
| 1  | If the consumer buys Chicken Eggs, then the consumer will also buy Beef Sausage Gaga | 10          | 100            | 10             |
| 2  | If a consumer buys the Beef Sausage Gaga, the consumer will also buy Chicken Eggs | 10          | 100            | 10             |
| 3  | If consumers buy Nescafe Coffee Cream 200ml, consumption will also buy Prima Water Mineral 600ml | 10          | 100            | 5              |
| 4  | If a consumer buys Bottle Sprite 1 Liter, then the consumer will also buy Pucuk Tea 350ml | 10          | 100            | 6.67           |
| No | Itemset                                                                                           | Support (%) | Confidence (%) | Lift Ratio (%) |
|----|---------------------------------------------------------------------------------------------------|-------------|----------------|---------------|
| 5  | If the consumer buys Aice Sweet Corn 52gr, then the consumer will also buy Dunhill Fine Cut Black Filter | 20          | 75             | 3.75          |
| 6  | If consumers buy Dunhill Fine Cut Filter Black, then consumers will also buy Aice Sweet Corn 52gr | 20          | 75             | 3.75          |
| 7  | If the consumer buys Aice Strawberry Crispy 55GR, then the consumer will also buy Aice Sweet Corn 52gr | 15          | 66.67          | 3.33          |
| 8  | If consumers buy Bear Brand 189ml, then consumers will also buy Aqua Bottle 1500ml                 | 15          | 66.67          | 2.67          |
| 9  | If the consumer buys Pucuk Tea 350ml, then the consumer will also buy Bottle Sprite 1 Liter         | 15          | 66.67          | 6.67          |
| 10 | If the consumer buys Roma Biscuit Kelapa 450gr, then the consumer will also buy Aqua Bottle 600ml   | 15          | 66.67          | 2.22          |

3.2. Result

Based on user observation and testing, the Consumer Purchase Pattern Information System shows that the application of the Apriori algorithm is very good to be implemented because it can produce a purchase pattern based on the transaction data used in this study. It is expected that the results of consumer purchasing patterns can help minimarket managers in making decisions to get even better profits.

4. Conclusion

The pattern of purchases by consumers obtained from the system will produce rules about the relevance of a product. The above analysis can be used for various purposes in sales, one of which is to regulate the placement of goods or the layout of goods and provide association rules by providing support for beverage values and confidence of beverage as guidance. The results by inserting the minimum confidence and support, that is if the level of the minimum confidence 65% and support is 10%. then, the resulting rules are as many as 10 rules. Based on the final results of the association rules in this study, there are 4 rules that have a value of 10% support and 100% confidence, namely: consumers buy Chicken Eggs, then consumers will buy Beef Sausage Gaga together, consumers buy Beef Sausage Gaga, consumers will buy Chicken Eggs together, consumers buy Nescafe Coffee Cream 200ml, consumers will buy 600ml of Prima Mineral Water together, consumers buy a 1 Liter Sprite Bottle, then consumers will buy The Pucuk 350ml.

References

[1] Surjandari I and Seruni A C 2010 MAKARA Technol. Ser. vol. 9 no. 2 pp. 1–5
[2] Valle M A, Ruiz G A and Morras R 2018 Expert Syst. Appl. vol. 97 pp. 146–162
[3] Gangurde R, Kumar D B and Gore D S D 2017 Building Prediction Model using Market Basket Analysis vol. 5 no. 2 pp. 1–8
[4] Videla-Cavieres I F and Rios S A 2014 Expert Syst. Appl. vol. 41 no. 4 pp. 1928–1936
[5] Putra P B I S and Suryani N P S M 2018 Int. J. Eng. Emerg. Technol. vol. 3 pp. 13–17
[6] Setiabudi D H, Budhi G S, Purnama I W J, and A Noertjahyana 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering Bali Indonesia pp. 196–199
[7] Chiang W ·Y 2018 Br. Food J. vol. 120 no. 3 pp. 665–675
[8] Dhanabhakyam D M and Punithavalli D M 2011 A Survey on Data Mining Algorithm for Market Basket Analysis pp. 24–28
[9] Mustakim et al. 2018 *J. Phys. Conf. Ser.* vol. 1114 pp. 1–9
[10] Setiawan A, Budhi G S, Setiabudi D H, and Djunaidy R 2017 *International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT)* Denpasar Bali Indonesia pp. 337–340
[11] Wen-Xiu X, Heng-nian Q, and H Mei-li 2010 *First International Conference on Networking and Distributed Computing* Hangzhou Zhejiang China pp. 309–313
[12] Chen M -C 2007 *Expert Syst. Appl.* vol. 33 no. 4 pp. 1110–1116
[13] Lam H Y, Ho G T S, Wu C H and Choy K L 2014 *Ind. Manag. Data Syst.* vol. 114 no. 5 pp. 711–733
[14] Septiani W, Marie I A, Sugiarito D and L Hakim 2019 *IOP Conf. Ser. Mater. Sci. Eng.* vol. 528 pp. 1–10
[15] Panjaitan S *et al.* 2019 *J. Phys. Conf. Ser.* vol. 1255 p. 012057
[16] Bhandari A, Gupta A and Das D 2015 *Procedia Comput. Sci.* vol. 46 pp. 644–651