Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns

Suprianto Panjaitan¹, Sulindawaty², Muhammad Amin³, Sri Lindawati⁴, Ronal Watrianthos⁵, Hengki Tamando Sihotang⁶, Bosker Sinaga⁷

¹,⁷ Teknik Informatika, STMIK Pelita Nusantara Medan, Sumatera Utara, Indonesia
²,⁶ Doctoral Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan-Indonesia
³ Program Doctoral Vocational Technology Education, Postgraduate Engineering Faculty Padang State University Padang West Sumatera -Indonesia
⁴ Manajemen Informatika, AMIK Logika, Sumatera Utara, Indonesia
⁵ Manajemen Informatika, AMIK Labuhan Batu, Sumatera Utara, Indonesia

* hengki_tamando@yahoo.com

Abstract. Consumer purchasing patterns are a form of purchases made by consumers, whether someone or a lot of people to get the desired item by making a purchase transaction. One characteristic of the purchase pattern is the existence of acquiring something through exchanging money. This study aims to create an application that is used in determining consumer purchasing patterns by applying a priori algorithms and using Visual Basic 2010 as a tool for determining consumer purchase patterns. This application uses a priori algorithm calculation method where the sample consumer purchase data will be sorted and calculated by providing the value of the minium support and configuration parameters and based on the results of confidence the largest number of conclusions such as: can be used as information for determining sales, the application of a priori algorithms can provide information pattern combination item set from consumer purchase data that is with support above 15% and confidence above 50% on item set.

1. Introducing

Cafe Bojack Coffee Shop is a business entity that is engaged in selling light dishes for the general public. where is this cafe one of hundreds of cafes in Medan City. Although the cafe is very crowded, but often experience problems such as not the availability of the consumer order menu even though the order is in demand. Basically, the Cafe owner of the Bojack Coffe Shop has not analyzed the data specifically for example to combine items which will cause problems so that they do not know how the relationship between items with other items. Data mining is the process of finding interesting patterns or information in selected data using certain techniques or methods. The techniques, methods, or algorithms in data mining vary greatly. The selection of the right method or algorithm depends very much on the objectives and the Knowledge Discovery in Database (KDD) process in its entirety. Data mining techniques to find associative rules or relationships between items are called association rule mining. one of the algorithms used to find association rules is a priori algorithm. The Apriori Algorithm, helps in forming possible combination item candidates, then tests whether the combination meets the minimum support parameters and minimum confidence which is the threshold value given by the user.
2. Research Methodology

2.1. Data Mining
Data mining or often referred to as knowledge discovery in databases (KDD) is an activity that includes gathering, using historical data to find order, patterns or relationships in large data. This data mining output can be used to help make future decisions. The development of KDD has caused the use of pattern recognition to diminish because it has become a part of data mining [4].

2.2. Association Rule
Following are the steps of association rules calculation process [3].

a. The system scans the database to get a 1-item set candidate (set of items consisting of 1 item) and calculates the support value. Then the support value is compared with the specified minimum support, if the value is greater or equal to the minimum support then the itemset is included in a large itemset.

b. Itemset which is not included in large itemsets is not included in the next iteration in pruning.

c. In the second iteration the system will use large itemset results in the first iteration (L1) to form the second itemset candidate (L2). In the next iteration, the system will use the results of large itemset in the iteration and then will use large itemset results in the previous iteration (Lk-1) to form the following itemset candidate (Lk). The system will join Lk-1 with Lk-1 to get Lk, as in the previous iteration the system will delete / prune the itemset combination which is not included in the large itemset.

d. After the join operation, the new itemset result from the join process is calculated for support.

e. The candidate formation process which consists of join and prune process will continue to be carried out until the candidate itemset set is null, or there are no more candidates to be formed.

f. After that, the result of frequent itemset was formed an association rule that met the specified support and confidence values.

g. In the formation of association rule, the same value is considered as one value.

h. The association rule that is formed must meet the specified minimum value.

Minimal Support: a value determined by the researcher to cut the combination of set items into fewer [2]. Minimal Confidence is a value that is also determined by the researcher to cut the combination of each k-item set (the result of minimal support trimming) to form association rules [2].

The basic methodology of association analysis is divided into two stages:

- **Support**
  Support from an association rule is the presentation of the combination of items in the database, where if have item A and item B then the support is the proportion of transactions in the database containing A and B [7].
  The support value of an item is obtained by the formula[6].

\[
\text{Support}(A) = \frac{(\text{The number of transactions containing } A)}{\text{Transaction Total}}
\]

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

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

\[
\text{Support}(A, B) = \frac{\sum \text{Transactions contain A and B}}{\sum \text{Transaction}}
\]

- **Confidence**
  Confidence of association rule is a measure of the accuracy of a rule, which is the presentation of a transaction in a database containing A and containing B [7].

\[
\text{Confidence} = P(B|A) = \frac{\sum \text{Transactions contain A and B}}{\sum \text{Transactions contain A}}
\]
2.3. Apriori Algorithm

Broadly speaking the a priori algorithm works [1].

a. The formation of candidate itemset, candidate k-itemset is formed from a combination of item (k-1) -itemset obtained from the previous iteration. One characteristic of the Apriori algorithm is the presence of candidate k-itemset trimming whose subset containing k-1 items is not included in the high frequency pattern with a length of k-1.

b. Calculation of support from each candidate k-itemset. Support from each of our candidate candidates is obtained by scanning the database to calculate the number of transactions that contain all items in the candidate items. This is also a feature of the Apriori algorithm which requires calculation by scanning the entire database as much as the longest item-itemet.

c. Set a high frequency pattern. A high frequency pattern that contains k items or itemset is determined from the candidate kitet whose support is greater than the minimum support.

d. If no new high frequency pattern is obtained, the whole process is stopped. If not, then k plus one and return to part 1.

3. Results and Discussion

3.1. Apriori Algorithm Analyst

a. High Frequency Pattern Analysis

This stage looks for item combinations that meet the minimum requirements of the support value in the database.

| Transaction | Items purchased                                      |
|-------------|------------------------------------------------------|
| 1           | Dark Chocolate, Sweet Tea, Fried Rice, Rice, Cow Eye Egg, Chicken Nugget |
| 2           | Sprite, Dark Chocolate, Cappuccino, Teh Tarik, Noodle soup, Fried noodles, Nutri Sari, Sweet Tea, Baked Banana |
| 3           | Fried Rice, Brown Chocolate, Dutch Eggplant, Sweet soup, Fried Chocolate, Fried Noodle, 
|             | Nutri Sari, Fried Rice, Fried Banana, Baked Banana   |
| 4           | Gravy Noodles, Fried Rice, Sweet Tea, Milk           |
| 5           | Teh Tarik, Kuku Bima, Sweet Tea, French Fries, Fried Noodle, Fried Rice |
| 6           | Milk, Fried Rice, Lemon Tea, Cappuccino, Aqua       |
| 7           | Fried Rice, Lemon Tea, Milo, Kuku Bima, Fried noodles |
| 8           | Cappuccino, Milo, Martabak Mie, Milk, Noodle soup   |
| 9           | French Fries, Milk Tea, Fried Rice, Nutri Sari, Cappuccino, Dark Chocolate |
| 10          | Sweet Tea, Cappuccino, Milo, French Fries, Fried Rice, Grilled Banana |
| 11          | Fried Noodles, Martabak Noodles, Noodles, Rice, Tarik Tea, Sweet Tea, Milk Tea |
| 12          | Noodle soup, Oranges, Rice, Cappuccino              |
| 13          | Nutri Sari, Cappuccino, Milo, Dutch Eggplant, Fried Rice, Coffee Mix, Avocado |
| 14          | Sweet Tea, Nasi Goreng, Coca-Cola, Baked Banana     |
| 15          | Nutri Sari, Teh Tarik, Cappuccino, Avocado, Fried noodles, Fried Rice, Noodle soup |
| 16          | Lemon Tea, Teh Tarik, Milk, Fried Rice, Noodle soup |
| 17          | Cappuccino, Sweet Tea, Milo, Noodles, Rice, Dark Chocolate |
| 18          | Dark Chocolate, Cappuccino, Fried Rice, Noodle soup |
| 19          | Fried Rice, Teh Tarik, Cappuccino, Fried Noodle, Dark Chocolate |

For example in this study the analyst wants a rule that has more than 15% support and more than 50% confidence.

Step 1. Look for K1 (1-itemset candidate) as follows.

| Item          | Candidate          |
|---------------|--------------------|
| Aqua          | 1/20 * 100% = 5%   |
| Avocado       | 2/20 * 100% = 10%  |
The International Conference on Computer Science and Applied Mathematic
IOP Conf. Series: Journal of Physics: Conf. Series 1255 (2019) 012057
doi:10.1088/1742-6596/1255/1/012057

| Candidate | Item          | Percentage |
|-----------|---------------|------------|
| 5%        | Brown Chocolate | 1/20 * 100% |
| 5%        | Coca-Cola      | 1/20 * 100% |
| 5%        | Chicken Nugget | 1/20 * 100% |
| 55%       | Cappuccino     | 11/20 * 100% |
| 35%       | Dark Chocolate | 7/20 * 100% |
| 15%       | French Fries   | 3/20 * 100% |
| 10%       | Kuku Bima      | 2/20 * 100% |
| 5%        | Coffee Mix     | 1/20 * 100% |
| 15%       | Lemon Tea      | 3/20 * 100% |
| 5%        | Fried noodles  | 7/20 * 100% |
| 45%       | Noodle soup    | 9/20 * 100% |
| 30%       | Milo           | 5/20 * 100% |
| 10%       | Martabak Noodles | 2/20 * 100% |
| 5%        | Eyes of Cattle Eggs | 1/20 * 100% |
| 70%       | Fried rice     | 14/20 * 100% |
| 20%       | Rice           | 4/20 * 100% |
| 20%       | Nutri Sari     | 4/20 * 100% |
| 5%        | Oranges        | 1/20 * 100% |
| 20%       | Grilled banana | 4/20 * 100% |
| 5%        | Fried bananas  | 1/20 * 100% |
| 10%       | Sprite         | 2/20 * 100% |
| 15%       | Milk           | 3/20 * 100% |
| 10%       | Dutch eggplant | 2/20 * 100% |
| 45%       | Sweet tea      | 9/20 * 100% |
| 10%       | Milk tea       | 2/20 * 100% |
| 35%       | Teh Tarik      | 7/20 * 100% |

The combination of itemset in f1 can be used as a 2-itemset broker. itemset-itemset from f1 that can be used are itemset-itemset which have similarities in the first k-1 item.

The 2-itemset candidate that can be formed from f1 is as follows.

| No | Item 1 | Item 2     | Candidate |
|----|--------|------------|-----------|
| 1  | Fried rice | Cappuccino | 7/20 * 100% = 35% |
| 2  | Fried rice | Noodle soup | 4/20 * 100% = 20% |
| 3  | Fried rice | Sweet tea  | 6/20 * 100% = 30% |
| 4  | Fried rice | Dark Chocolate | 5/20 * 100% = 25% |
| 5  | Fried rice | Teh Tarik  | 4/20 * 100% = 20% |
| 6  | Fried rice | Fried noodles | 5/20 * 100% = 25% |
| 7  | Fried rice | Milo       | 3/20 * 100% = 15% |
| 8  | Fried rice | White rice | 1/20 * 100% = 5% |
| 9  | Fried rice | Nutri Sari | 3/20 * 100% = 15% |
| 10 | Fried rice | Grilled banana | 2/20 * 100% = 10% |
| 11 | Fried rice | milk       | 3/20 * 100% = 15% |
| 12 | Nasi Goreng | Lemon Tea  | 3/20 * 100% = 15% |
| 13 | Fried rice | French Fries | 3/20 * 100% = 15% |
| 14 | Cappuccino | Noodle soup | 6/20 * 100% = 30% |
| 15 | Cappuccino | Sweet tea  | 3/20 * 100% = 15% |
| 16 | Cappuccino | Dark Chocolate | 5/20 * 100% = 25% |
| 17 | Cappuccino | Teh tarik  | 3/20 * 100% = 15% |
| 18 | Cappuccino | Fried noodles | 3/20 * 100% = 15% |
| 19 | Cappuccino | Milo       | 4/20 * 100% = 20% |
| 20 | Cappuccino | rice       | 2/20 * 100% = 10% |
| 21 | Cappuccino | Nutri Sari | 4/20 * 100% = 20% |
| 22 | Cappuccino | Grilled banana | 2/20 * 100% = 10% |
| No  | Item 1  | Item 2        | Candidate        |
|-----|---------|--------------|------------------|
| 23  | Cappucino | White milk   | 2/20 * 100% = 10% |
| 24  | Cappucino | Lemon Tea    | 1/20 * 100% = 5%  |
| 25  | Cappucino | French Friesh| 2/20 * 100% = 10% |
| 26  | Noodle soup | Sweet tea   | 4/20 * 100% = 20% |
| 27  | Noodle soup | Dark Chocolate | 3/20 * 100% = 15% |
| 28  | Noodle soup | Teh Tarik  | 4/20 * 100% = 20% |
| 29  | Noodle soup | Fried noodles | 3/20 * 100% = 15% |
| 30  | Noodle soup | Milo        | 2/20 * 100% = 10% |
| 31  | Noodle soup | White rice  | 3/20 * 100% = 15% |
| 32  | Noodle soup | Nutri Sari  | 2/20 * 100% = 10% |
| 33  | Noodle soup | Grilled banana | 1/20 * 100% = 5% |
| 34  | Noodle soup | milk        | 3/20 * 100% = 15% |
| 35  | Noodle soup | Lemon Tea   | 1/20 * 100% = 5%  |
| 36  | Noodle soup | French Fries | 0/20 * 100% = 0%  |
| 37  | Sweet tea  | Dark Chocolate | 4/20 * 100% = 20% |
| 38  | Sweet tea  | Teh Tarik  | 3/20 * 100% = 15% |
| 39  | Sweet tea  | Fried noodles | 4/20 * 100% = 20% |
| 40  | Sweet tea  | Milo        | 2/20 * 100% = 10% |
| 41  | Sweet tea  | rice        | 3/20 * 100% = 15% |
| 42  | Sweet tea  | Nutri Sari  | 1/20 * 100% = 5%  |
| 43  | Sweet tea  | Grilled banana | 3/20 * 100% = 15% |
| 44  | Sweet tea  | milk        | 1/20 * 100% = 5%  |
| 45  | Sweet tea  | Lemon Tea   | 0/20 * 100% = 0%  |
| 46  | Sweet tea  | French Fries | 2/20 * 100% = 10% |
| 47  | Dark Chocolate | Teh Tarik | 2/20 * 100% = 10% |
| 48  | Dark Chocolate | Fried noodles | 3/20 * 100% = 15% |
| 49  | Dark Chocolate | Milo   | 1/20 * 100% = 5%  |
| 50  | Dark Chocolate | White rice | 2/20 * 100% = 10% |
| 51  | Dark Chocolate | Nutri Sari | 2/20 * 100% = 10% |
| 52  | Dark Chocolate | Grilled banana | 1/20 * 100% = 5% |
| 53  | Dark Chocolate | milk    | 0/20 * 100% = 0%  |
| 54  | Dark Chocolate | Lemon Tea | 0/20 * 100% = 0%  |
| 55  | Dark Chocolate | French Fries | 1/20 * 100% = 5% |
| 56  | Teh Tarik  | Milo        | 1/20 * 100% = 5%  |
| 57  | Teh Tarik  | White rice  | 1/20 * 100% = 5%  |
| 58  | Teh Tarik  | Nutri Sari  | 2/20 * 100% = 10% |
| 59  | Teh Tarik  | Grilled banana | 2/20 * 100% = 10% |
| 60  | Teh Tarik  | milk        | 1/20 * 100% = 5%  |
| 61  | Teh Tarik  | Lemon Tea   | 1/20 * 100% = 5%  |
| 62  | Teh Tarik  | French Friesh | 1/20 * 100% = 5% |
| 63  | Fried noodles | Milo   | 1/20 * 100% = 5%  |
| 64  | Fried noodles | rice    | 1/20 * 100% = 5%  |
| 65  | Fried noodles | Nutri Sari | 2/20 * 100% = 10% |
| 66  | Fried noodles | Grilled banana | 1/20 * 100% = 5% |
| 67  | Fried noodles | milk    | 0/20 * 100% = 0%  |
| 68  | Fried noodles | Lemon Tea | 1/20 * 100% = 5%  |
| 69  | Fried noodles | French Fries | 1/20 * 100% = 5% |
| 70  | Milo       | White rice  | 1/20 * 100% = 5%  |
| 71  | Milo       | Nutri Sari  | 1/20 * 100% = 5%  |
| 72  | Milo       | Grilled banana | 2/20 * 100% = 10% |
| 73  | Milo       | milk        | 1/20 * 100% = 5%  |
| 74  | Milo       | Lemon Tea   | 1/20 * 100% = 5%  |
| 75  | Milo       | French Fries | 0/20 * 100% = 0% |
| 76  | Rice       | Nutri Sari  | 0/20 * 100% = 0%  |
| 77  | Rice       | Grilled banana | 0/20 * 100% = 0% |
| 78  | Rice       | milk        | 0/20 * 100% = 0%  |
After all high frequency patterns have been found, then the association rules that meet the minimum requirements are found for confidence associative rules $A \rightarrow B$ minimum confidence = 50%.

Table 4. Candidate association rules from F2

| Itemset | Support | Confidence |
|---------|---------|------------|
| If you buy Fried Rice, will buy Cappuccino | 35% | 50% |
| If you buy Cappuccino, you will buy Noodles soup | 30% | 54.55% |
| If you buy Fried Rice, will buy Rice | 30% | 54.55% |
| If you buy Fried Rice, will buy Teh Tarik | 53% | 71.43% |
| If you buy Teh Tarik, will buy Fried Noodles | 53% | 71.43% |

b. Final Association Rules

Association rules are sequential based on minimum support and minimum confidence that meet the following:

Table 5. Candidate association rules from F2

| Itemset | Support | Confidence |
|---------|---------|------------|
| If you buy Fried Rice, you will buy Cappuccino | 35% | 50% |
| If you buy Cappuccino, you will buy Noodles soup | 30% | 54.55% |
| If you buy Teh Tarik, you will buy Fried Noodles | 25% | 71.43% |

c. System Design
3.2. Implementation System
This page is used to display the results of the data mining process.

4. Conclusions
Based on the discussion and implementation, several conclusions can be obtained:
1. Analysis of consumer purchasing patterns using a priori algorithm in this application, can provide association rules by giving the drink value support and drink confidence as a reference.
2. From the results of testing by entering a different value of support and drinking confidence, that is, if the minimum support is 15% and the confidence level is 50%, a rule of 87 rules will be generated.
3. The benefits of the test results for the cafe include:
   a. Fried Rice Stocks and Cappucino stock can be provided with many of the same stock.
   b. Cappucino stock and stock of Kuah Noodle can be provided with many of the same stock.
   c. The tea stock pull and stock of fried noodles can be provided with many of the same stock.
   d. The cafe can make a jumbo package food menu for more interesting

References
[1] Achmad Nur Cahyon, "Implementation of Data Mining of Apriori Algorithms on Motorcycle Spare Parts Sales in Ahas Putra Motor", 2016
[2] Al-maolegi, Mohammed, and Bassam Arkok. A N I MPROVED A PRIORI A LGORITHM
FOR. Vol. 3, no. 1, 2014, pp. 21–29.

[3] Bhandari, Akshita, et al. “Improvised Apriori Algorithm Using Frequent Pattern Tree for Real Time Applications in Data Mining.” Procedia - Procedia Computer Science, vol. 46, no. Icict 2014, Elsevier Masson SAS, 2015, pp. 644–51, doi:10.1016/j.procs.2015.02.115.

[4] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. P. Windarto, “Application of Classification Method C4.5 on Selection of Exemplary Teachers,” in IOP Conference Series, 2018, pp. 1–6.

[5] Desti Fitriati, "Implementation of Data Mining to Determine the Combination of Goods Promotion Media Based on Customer Purchase Behavior Using Apriori Algorithms", Annual Research Seminar, 2016, 2 (1)

[6] Domna Lingga, "Application of Apriori Algorithms in Predicting Book Inventory at the Dwi Tunggal Tanjung Morawa High School Library", Information and scientific technology, 2016, XI (1)

[7] Dutt, Shalini, et al. “An Improved Apriori Algorithm Based on Matrix Data Structure.” Global Journal of Computer Science and Technology: C Software & Data Engineering, vol. 14, no. 5, 2014, pp. 1–5.

[8] Fadlina "Data Mining For Analysis of Street Crime Levels with Apriori Association Rule Algorithm Apriori Method", Scientific Information and Technology (CORE), 2014, III (1)

[9] S. Sudirman, A. P. Windarto, and A. Wanto, “Data Mining Tools | RapidMiner: K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia,” IOP Conference Series: Materials Science and Engineering, vol. 420, no. 12089, pp. 1–8, 2018.

[10] Liu, Xiyu, et al. “An Improved Apriori Algorithm Based on an Evolution-Communication Tissue-Like P System with Promoters and Inhibitors.” Hindawi Discrete Dynamics in Nature and Society, vol. 2017, 2017, pp. 1–12.

[11] Mohammad Fauzy, Pack of Rahmat Saleh W, Ibnu Asror, "Application of the Rule Association Using the Apriori Algorithm in Rain Prediction Simulation in Bandung City Region,” Applied Scientific Information Journal, 2016.

[12] Palekar, Akshay, et al. “Data Mining Using Apriori Algorithm.” International Journal of Engineering Trends and Technology (IJETT), vol. 28, no. 4, 2015, pp. 190–92.

[13] Suresh, Jogi, and T. Ramanjaneyulu. “Mining Frequent Itemsets Using Apriori Algorithm.” International Journal of Computer Trends and Technology (IJCTT), vol. 4, no. April, 2013, pp. 760–64.

[14] Yotenko, Rahmadi. “The Implementation Of Apriori Algorithm And Chi-Square Test In Determining Pattern Of Relationship Among The Rawi Hadis.” Eksakta: Jurnal Imu-Ilmu MIPA, vol. 17, no. 1, 2017, pp. 20–32.

[15] Yuan, Xiuli. “An Improved Apriori Algorithm for Mining Association Rules.” Advances in Materials, Machinery, Electronics I Conference, vol. 080005, no. 2017, 2018, doi:10.1063/1.4977361.