Implementation and comparison analysis of apriori and fp-growth algorithm performance to determine market basket analysis in Breiliant shop

A A Hidayat*, A Rahman, R M Wangi, R J Abidin, R S Fuadi and W Budiawan
Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia

*acep@if.uinsgd.ac.id

Abstract. The creativity of an online store owner who develops a business in a retail business in determining a marketing strategy will affect his competitive ability with other online stores. However, services that are in accordance with costumers attitude is a stuff that should be awarded by shop owners in order to improve not only customer satisfaction but also store income. The method used in this study is Market Basket Analysis to know customers attitude by analysing data from sales transactions to show customers attitude toward one sold item and the others. Research aimed that it is necessary to have a system that can help to form a combination among products using the association rule method, the algorithm used to analyse this method is apriori and fp-growth algorithm. Result from data analysis test using monthly sales transaction data of cosmetics in Breiliant Store during November 2018 with 34 number of data sales transaction. It could be concluded that a combination of products which have strong support and confidence were Original Liquid Bleaching Seeds, Harva Peeling Gel and Castor oil. It had support value of 8.8% and 30% confidence value, with a filtering time of 0.036 seconds.

1. Introduction

The development and competition in the world of trading business and the advancement of information technology is a matter that is interrelated, in the tight competition of the market to meet the increasingly high demands of customers. Companies need business strategy and intelligence to continue to meet customer desires and market demands. Regardless of the type of market involved, business people are competing to provide the best service, it is viewed from the vital aspects that become material for consumer assessment.

Having its origin in the analysis of the marketing bucket, the exploration of association rules represents one of the main applications of data mining. Their popularity is based on an efficient data processing by means of algorithms. Being given a set of transactions of the clients, the purpose of the association rules is to find correlations between the sold articles. Knowing the associations between the offered products and services, helps those who have to take decisions to implement successful marketing techniques [1].

The apriori is first algorithm to generate all frequent item sets and confident association rules was the AIS algorithm by Agrawal [1-3]. The first process uses the Apriori algorithm to determine the frequent sets and to generate association rules based on the frequent sets discovered [1,4].

FP growth algorithmic program is an efficient algorithm for producing the frequent item sets without generation of candidate item sets. It adopts a divide and conquer strategy and it needs two database scans
to seek out the Support count. It can mine the items by using lift, leverage and conviction by specifying minimum threshold [3,5]. The FP-Growth algorithm is a development of Apriori, the deficiency of the Apriori algorithm improved by the FP-Growth algorithm [6].

Association rule of data mining involves preference out the nameless inter-relation of the data and finding out the rules between individual items [7], mining for association rules between items in a large database of sales transactions has been described as an important database mining problem [8]. Data mining is from a lot of, incomplete, noisy, fuzzy and random data, extract implicit in them, people don't know in advance, but it is potential useful information and knowledge of the process. Association rule mining is from a database search item sets between meaningful connection or related links. It is an important technology in data mining, in recent years, it is widely used in various fields [9].

Breiliant shop is a business that develops its business in the field of beauty. The store sells various kinds of cosmetic products, skin care, hair care and so on. The Breiliant shop is an online shop that has been joined for approximately 3 years with the Shopee marketplace.

The more mushrooming of shops that also sell similar products, the competition also increases for it, so strategies are needed to maintain the business. In connection with this, producers must be able to understand what consumers really want to provide comfort in shopping at the store, especially in making it easy to choose the groceries desired by consumers easily. For example, in laying out groceries arranged in the application homepage, it should be adjusted to consumer spending patterns to make it easier for consumers to find the desired items.

The method often used to analyse consumer shopping behaviour patterns is shopping basket analysis or Market Basket Analysis (MBA). MBA is one of the most popular types of data analysis used in the marketing world [10]. The purpose of Market Basket Analysis is to determine what products are most commonly purchased or used by consumers [6]. This MBA is analysing consumer buying habits by finding associations between different products that consumers place in shopping basketball [11]. In general, MBA is an example of the implementation of the Association Rule [6]. This analysis is one method in data mining that learns about consumer behaviour in buying goods simultaneously at one time. Many methods are used in data mining, namely estimation, prediction, classification, clustering, Association. This study uses the association rule using the Apriori algorithm and FP Growth. Both of these algorithms are used to determine the correlation between items that are in demand by consumers stored in the database. After obtaining frequent item sets, a rule is taken and then analyse the comparison of the time of the two algorithms

2. Research methods

Consumer behaviour is the things that underlie consumers to make purchasing decisions. AMA (American Marketing Association) defines consumer behaviour as a dynamic interaction between cognition, affection, behaviour and the environment in which a person engages in exchange activities in their lives. From here can be taken 3 important things, namely:

- Consumer behaviour is dynamic, so it is difficult to predict.
- Involving interactions, such as cognition, affection, behaviour and events around consumers.
- Involving exchanges, such as the exchange of goods and money from the seller to the buyer.

Association rules are certain rules or rules that state a correlation between the levels of occurrence of several attributes in a database. The general forms of Association Rules are:

\[ A_1 \ldots A_n \rightarrow B \] (1)

Which means that customers who buy product A also have a considerable opportunity to buy product B, where there is no limit on the number of items on the heal or body of a rule. The importance of an associative rule can be known by two parameters, namely support and confidence. Support
value) is the percentage of the combination of items in the database. While Confidence (the value of certainty) is the strong relationship between items in the rules of association.

The support value of an item is obtained by calculation as follows.

\[
Support(A \cap B) = \frac{\text{number of transactions containing } A \text{ and } B}{\text{total number of transactions overall}} \times 100\%
\]  

(2)

\[
Confidence(A \mid B) = \frac{\text{number of transactions containing } A \text{ and } B}{\text{number of transactions containing } A} \times 100\%
\]  

(3)

3. Result and discussions

Based on the existing sales transaction data then the sales trend analysis is carried out to find out the product combination in the Breiliant store, the data analysed is the cosmetic sales transaction data in November 2018. The data is data that represents 34 transaction data as a whole, can be seen in Table 1

| Transaction | Items purchased |
|-------------|-----------------|
| 1           | Bibit pemutih cair ori, Harva peeling gel, Lotion Vampir 500 ml, Miss Moter hand wax |
| 2           | Castor oil minyak jarak, Lipstik flower crysantium, Egg white mask Serum vampire, Nose up pemancung hidung |
| 3           | Bibit pemutih cair ori, Bedak Implora, Eyeliner Pencil, Blackhead remover |
| 4           | Vaseline lip therapy, Pencil alis loreal, Baby Pinkiss, Cream HN kecil |

3.1. Grouping transaction data

Transactions that existed in November 2018 were 34 transactions with a total of 126 products in the transaction. In each transaction there are the same types of items so that after grouping the results are obtained, there are 47 types of product items as seen in the table 2.

| No | Itemset | Item Amount |
|----|---------|-------------|
| 1  | Bibit pemutih cair ori | 9 |
| 2  | Harva peeling gel | 9 |
| 3  | Lotion Vampir 500 ml | 2 |
| 4  | Miss Moter hand wax | 4 |
| 5  | Castor oil minyak jarak | 7 |

3.2. High frequency pattern search analysis

Pattern search analysis is done to find associations using a priori algorithm by entering the minimum support and minimum confidence values, where the greater the minimum support value that is included is the more effective the results of the rules of the resulting product combination. The data tested consisted of 34 sales transactions in which there were 47 different types of beauty products so that the total number of products contained in the transaction was 126 products. The test results on testing using a priori algorithm can be seen in the table 3.

| Association Rules                          | Support | Confidence | Execution Time       |
|-------------------------------------------|---------|------------|----------------------|
| Bibit pemutih cair ori, Harva peeling gel | 8.8 %   | 33%        | 1 hour 34 seconds    |
| Castor oil minyak jarak                   |         |            |                      |
The characteristic of the Fp-growth algorithm is that the data structure used is a tree called FP-Tree, the Fp-growth algorithm can extract Frequent Itemset from FP-Tree. Frequent excavation of itemset using the Fp-growth algorithm will be carried out by generating a data tree structure, called FP-Tree. The Fp-growth method can be divided into 3 main stages, namely:

- Phase generation conditional pattern base
- Stage of conditional FP-Tree generation, and
- Search phase for Frequent Itemset.

Steps to finding rules with Fp-growth

3.2.1. Generate frequent itemset. For example, Min has a 30% Support or a minimum of 3 transactions, then a product id that is less than 3 transactions is eliminated. Then the data that meets the minimum support is sorted by frequency, the table is usually called FP-List.

Table 4. Formation of FP-List.

| No | Product ID | Occurrence | Frequency |
|----|------------|------------|-----------|
| 1  | 82         | 9          |
| 2  | 66         | 9          |
| 3  | 35         | 7          |
| 4  | 70         | 5          |
| 5  | 58         | 5          |
| 6  | 38         | 5          |
| 7  | 88         | 5          |
| 8  | 63         | 5          |
| 9  | 5          | 4          |
| 10 | 32         | 4          |
| 11 | 67         | 4          |
| 12 | 57         | 4          |
| 13 | 76         | 4          |
| 14 | 9          | 4          |
| 15 | 17         | 3          |
| 16 | 52         | 3          |
| 17 | 15         | 3          |
| 18 | 59         | 3          |

3.2.2. Add Transaction ID (TID) to the dataset that has been selected with minimum support. The function of this TID is to give a sequence number to the transaction after creating an F-List. Sort items in each transaction based on the highest to lowest frequency. Then start forming the Tree in sequence based on the Transaction ID (TID).

Table 5. Sorting datasets.

| TID | Transaction dataset | Selected datasets |
|-----|---------------------|-------------------|
| 1   | { 82, 66, 14, 15 }  | 82, 66, 15        |
| 2   | { 35, 32, 20, 17, 65 } | 35, 32, 17       |
| 3   | { 82, 6,78, 70 }    | 82, 70            |
| 4   | { 30, 93, 67, 62 }  | 67                |
| 5   | { 23, 54, 57 }      | 57                |
| 6   | { 58, 38, 35, 2 }   | 35, 58, 38        |
| 7   | { 35, 66, 57, 12, 72 } | 66, 35, 57, 72   |
| 8   | { 88, 76, 35, 32, 14 } | 35, 88, 76, 32   |
| 9   | { 38, 59, 63 }      | 38, 63, 59        |
Table 5. Cont.

| TID | Transaction dataset | Selected datasets |
|-----|----------------------|------------------|
| 10  | { 5, 15, 9, 19 }     | 5, 9, 15         |
| 11  | { 57, 63, 82, 35, 88 } | 82, 35, 63, 88, 57 |
| 12  | { 66, 23, 67, 8, 35 } | 66, 35, 67       |
| 13  | { 88, 87, 5, 38 }    | 88, 5, 38        |
| 14  | { 83, 90 }           | -                |
| 15  | { 91, 96, 88, 32 }   | 88, 32           |
| 16  | { 70, 19, 9 }        | 70, 9            |
| 17  | { 38, 82, 58 }       | 83, 38, 58       |
| 18  | { 65, 76, 81, 90, 1 } | 65, 76          |
| 19  | { 38, 59 }           | 38, 59           |
| 20  | { 70, 9, 72, 17 }    | 70, 9, 72, 17    |
| 21  | { 15, 20, 44, 51 }   | 15               |
| 22  | { 61, 15, 24, 72, 66 } | 66, 15, 72     |
| 23  | { 18, 66, 1, 63 }    | 66, 63           |
| 24  | { 70, 12 }           | 70               |
| 25  | { 35, 32, 52, 17, 9 } | 35, 32, 9, 52, 17 |

3.2.3. Form a Frequent Pattern tree (FP-Tree)

FP-Tree is a compressed data storage structure. FP-Tree is built by mapping each transaction data into a certain path in the FP-Tree, because in every transaction that is mapped, there may be transactions that have the same item, then the trajectory allows it to overlap each other.

The more transaction data that has the same item, the compression process with the FP-Tree structure is more effective. The advantage of FP-Tree is that it requires twice the scanning of transaction data that is produced very efficiently.

The FP-Tree is a tree with the following definition:

- FP-Tree is formed by a root labelled Null, a group of trees consisting of certain items and a frequent header table.
- Each node in FP-Tree contains three important information, namely the item label, informing the type of item represented by the node. Support count, presents the number of transaction paths that pass through that node, and the connecting pointer that connects the nodes with the same item label.

Then the test results are obtained as follows.

Table 6. Test results for FP-Growth algorithms.

| Association Rules | Support | Confidence | Execution Time |
|-------------------|---------|------------|----------------|
| Bibit pemutih cair ori, Harva peeling gel ➔ Castor oil minyak jarak | 8.8 % | 30% | 0.036 seconds |

4. Conclusions

Based on the discussion, implementation and testing of the system, some conclusions are obtained:

- The Association Rule method using the Priori Algorithm and FP-Growth Algorithm with parameters of support and confidence can obtain a correlation of purchase items to further improve the sales strategy and in terms of promotion of goods in order to increase the revenue for shop owners.
- A combination of products that have strong support and confidence, namely liquid ori whitening seeds, Harva peeling gel and Castor oil castor oil. Time needed The FP-Growth algorithm is
faster when compared to the Apriori Algorithm in obtaining the results of correlation / product combination.

- In the process the a priori algorithm uses the name of the product / product and produces high rules of accuracy value while the fp-growth algorithm uses product / product codes with rules that are lower than the priori algorithm.

For this case, because we use implementation carried out using a combination of a priori algorithms with Fp Growth for this case it is considered a priori better than data accuracy. It is also possible for other cases to have better growth. In the next study, it is expected to use larger data by selecting combinations using other algorithms, so that it will produce more rules as well.

Acknowledgments
We would like to say thank you for Research and Publication Centre of UIN Sunan Gunung Djati Bandung that give the full support and fund for this publication research.

References
[1] Hunyadi D 2011 Performance comparison of Apriori and FP-Growth algorithms in generating association rules Proc. Eur. Comput. Conf. 376–381
[2] Agrawal R, Imielinski T and Swami A 1993 Sigmod93 Assoc Min. Assoc. Rules between Sets Large Database 1–10
[3] S Mythili M and R Mohamed Shanavas A 2013 Performance Evaluation of Apriori and FP-Growth Algorithms Int. J. Comput. Appl. 79 34–37
[4] Zarman A, Irfan M and Uriawan W 2016 Implementasi Algoritma Ant Colony Optimization Pada Aplikasi Jurnal 1 6–12
[5] Han J and Kamber M 2001 Data mining:concept and techniques
[6] Rahim R, Manulangga G, Kurniasih N, Daeng G S A, Mustakim, Iskandar A, Herianda D M, Laumal F E, Ilham A and Indra Iswara I B A 2018 Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau J. Phys. Conf. Ser. 1114 012131
[7] Singh A K, Kumar A and Maurya A K 2015 An empirical analysis and comparison of apriori and FP-growth algorithm for frequent pattern mining Proc. 2014 IEEE Int. Conf. Adv. Commun. Control Comput. Technol. ICACCCT 1599–1602
[8] Omiecinski E, Savasere A and Navathe S 1995 An Efficient Algorithm for Mining Databases Association Rules in Large Proc. 21st VLDB Conf. 432–444
[9] Lei G, Dai M, Tan Z and Wang Y 2011 The research of CMMB wireless network analysis based on data mining association rule 7th Int. Conf. Wirel. Commun. Netw. Mob. Comput. WiCOM 2011 11–14
[10] Mujtaba A 2007 Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management 67
[11] Deshmukh D and More A 2017 International Journal of Innovative Research in Computer and Communication Engineering Applying Big Data in Higher Education