Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau

Mustakim¹, Della Maulina Herianda², Ahmad Ilham³, Achmad Daeng GS⁴, Folkes E. Laumal⁵, Nuning Kurniasih⁶, Akbar Iskandar⁷, Gloria Manulanga⁸, Ida Bagus Ary Indra Iswara⁹ and Robbi Rahim¹⁰

¹Puzzle Research Data Technology, UIN Sultan Syarif Kasim, Riau, Indonesia
²Department of Information System, UIN Sultan Syarif Kasim, Riau, Indonesia
³Informatics Department, Universitas Muhammadiyah Semarang, Semarang, Indonesia
⁴Department of Management, University of 45 Surabaya, Surabaya, Indonesia
⁵Department of Electrical Engineering, Politeknik Negeri Kupang, Kupang, Indonesia
⁶Faculty of Communication Sciences, Library and Information Science Program, Universitas Padjadjaran, Bandung, Indonesia
⁷Informatics Engineering, STMIK AKBA, Makassar, Indonesia
⁸Department of Electrical Engineering, Politeknik Negeri Kupang, Kupang, Indonesia
⁹Informatics Department, STMIK STIKOM Indonesia, Denpasar, Indonesia
¹⁰School of Computer and Communication Engineering, Universiti Malaysia Perlis, Kubang Gajah, Malaysia

*mustakim@uin-suska.ac.id

Abstract. Berkah Mart is one of the new minimarkets in Pekanbaru city which began to develop as a retail business to meet the needs of the community. In his operation, Berkah Mart still uses the traditional store management perception that the placement of product layout on sale shelves has not been through sales analysis. Existing products have not been reviewed from the consumer's point of view, whereas the marketing economics concept provides a variety of product and consumer management methods for business management in Berkah Mart, including market bucket analysis techniques and various itemset search algorithms. In this research, Market Basket Basket Analysis with FP-Growth algorithm is proposed to determine the layout and planning of goods availability. The application of FP-Growth algorithm proved to be useful in generating many and informative association rules to find out the consumer spending pattern at Berkah Mart in Pekanbaru. Moreover, from the rules of the customer association can be segmented separately to meet the specific needs of customers with cost-effective by using some special promotions for the general group. The experimental results show that FP-Growth algorithm can analyze quickly and efficiently informing consumer shopping pattern at Berkah Mart and can increase Berkah Mart revenue.
1. Introduction

In recent years, the growth of retail business in Pekanbaru City has increased quite high. This growth can be seen from the operation of several modern stores such as Minimarket, Supermarket, Department Store and Hypermart provide various community needs. The increasing number of retail business is influenced by the increasing demand and the society's need for goods as the necessities of life [1].

On the one hand, there are many modern stores in Pekanbaru where it gives its own pride as a provincial city that shows the progress of the city, but in terms of business management, the presence of a modern store gives the effect of competition among retail business companies as it provides almost the same community needs. Therefore, the management company should have and develop a variety of management strategies and increase sales/marketing is appropriate to be able to stay afloat in business and still have consumers. One strategy that can be undertaken by the management of a modern store business is to control the movement of products and patterns of consumers to various products sold using various business methods applicable in the field of marketing management science [2]. Management strategies that may impact risk include product selection, in-store product layout, quantity and quality of products sold, availability of other products that are mutually correlated, service ethics, safety, convenience and product pricing [3].

Berkah Mart is one of the new mini markets in Pekanbaru city which also move in retail business to fulfill public goods purchase. As new stores begin to evolve the various risks of business competition can occur. In its operation, Berkah Mart has implemented a service that is quite interesting in terms of availability of product information, security, comfort and employees are friendly to consumers. However, the placement of product layout on the sales shelves is still less accurate and optimal because it is done only based on store management perception by categorizing existing products and has not been reviewed from the point of view of consumers, whereas the concept of marketing economics provides various methods of management of products and consumers to business management in Berkah Mart, such as bucket analysis model mock and search algorithm frequent item set. This research has analyzed data by using Market Basket Analysis technique with association rule algorithm to determine the layout and planning of goods availability. This study compares the two algorithms associated with Apriori and Frequent Pattern Growth (FP-Growth). From both algorithms there are two wanted things to know from dataset blessing mart which is rule process that resulted from each algorithm and time of achievement of algorithm in forming a rule.

The data involved in any sale transaction on this Berkah Mart, such as data of items purchased, time of purchase, total sales volume, item price. Berkah Mart requires additional data for managers to make strategic decisions that can increase company profits, such as the most sold product information, slightly sold products, and rarely sold products. Maintain inventory is to know the pattern of consumer spending that often occurs at Berkah Mart by analyzing the data of sales transactions that amounted to more than 400 transactions/day. The placement of the product layout is still less accurate and optimal because it is only based on management’s perception by categorizing the existing products and has not been reviewed from the consumer's point of view. So that, the researcher’s initiative to try to provide solutions in the placement of the product layout.

The association pattern obtained from the proposed method will be used as a reference in determining the neat arrangement of items according to the combination of most frequently purchased and interconnected items. The right association pattern in every product laying can improve service and sale of goods to consumers with appropriate marketing strategy [4]. In this case study investigated 3 months of transactions ie in October, November, December. Association rule mining is applied using 7 experiments, namely October (1), November (2), December (3), October and November (4), October and December (5), November and December (6), October, November and December (7). The purpose of this experiment is to further reinforce the end result of any combination of Association rule mining.
2. Literature review

2.1. Berkah mart
Berkah Mart is a retail company that has been established since 2014 and rescued on Jl. Like Work/ Kualu, Panam, Pekanbaru. This minimarket has a good business process that is the number of stable consumers in the period of each month. The products are often given discounts are the basic ingredients like other minimarkets. Berkah Mart Memiliki Some competitors such as Indomaret, Alfamart, Planet Supermarkets are not too far away with Berkah Mart. Based on interviews and field observations, Berkah Mart has 8,307 items of goods sold there in December 2017, where the number of transactions that come out in a day average of 400 transactions, in a single transaction there is 1 item and maximum there are 20 items [5]. Layout is very influential on consumer purchases, due to the number of consumers who ask the employees where they want the goods.

2.2. Market based analysis (MBA)
MBA is one of the most popular types of data analysis used in the marketing world [6]. The purpose of Market Basket Analysis is to determine what products are most commonly purchased or used by consumers [7]. This MBA is analyzing consumer buying habits by finding associations between different products that consumers place in shopping basket [8]. In general, MBA is one example of the implementation of the Association Rule

2.3. Association rule mining (ARM)
ARM is a procedure for finding relationships between items of a defined dataset [9] and searching for and finding relationships between items in a dataset [10] Implementation of data mining with the rules of association aims to find information of items that are alternating since in the form of rules. The rules of association are data mining techniques for finding association rules among a combination of items [11]–[13]

2.4. Apriori and FP-Growth Algorithm
The Apriori Algorithm is a basic algorithm proposed by Agrawal & Srikant in 1994 for the determination of the frequent itemset for boolean association rules. A priori algorithm includes the type of association rules in data mining. The rule that states associations between multiple attributes is often called affinity analysis or market basket analysis [10], [11]. The a priori algorithm is divided into several stages called narrative which consists of (1) The establishment of candidate items, (2) Calculation of support of each k-itemset candidate, (3) Set the high frequency pattern and (4) If no new high frequency pattern is obtained the whole process is stopped. If not, then k plus one and return part 1 [14].

The FP-Growth algorithm is a development of Apriori, the deficiency of the Apriori algorithm improved by the FP-Growth algorithm [15]. In Apriori a generate candidate is required to get frequent itemsets. However FP-Growth generate candidate algorithm is not done because FP-Growth uses the concept of tree development in search of the frequent itemsets. This is what causes the FP-Growth algorithm is faster than the Apriori algorithm [16].

3. Proposed method
This research started with data collection from Berkah Mart Pekanbaru by experiment modeling is 7 experiments. Furthermore, the Apriori and FP-Growth algorithm processes are compared based on the combination of each item set rule and the time efficiency comparison used in establishing a rule. The comparison result of both algorithms obtained an efficient algorithm to be applied at Berkah Mart Pekanbaru as a material of the analysis and the improvement of sales strategy.
4. Analysis and results

Based on the research methodology, several important things that will be done to complete this research consist of planning, data collection and processing, Association Rule calculation using a priori algorithm, and data validation using Weka 3.7.4 and Tanagra 1.4. Table 1 shown Berkah Mart dataset.

Table 1. Berkah mart dataset.

| Range Data | Monthly Transaction | Daily Transaction | Goods out month | Goods out period |
|------------|---------------------|------------------|-----------------|-----------------|
| October    | 12                  | 400              | 74.586          | 2.406           |
| November   | 15.6                | 520              | 85.380          | 2.846           |
| December   | 14.4                | 4.80             | 82.460          | 2.660           |

4.1. Data collection

The data used has been done preprocessing, after this doing a combination of experiments. Experiments were performed seven times in the data: October, November, December, October-November, October-December, November-December, October-November-December. In the process of analyzing this a priori algorithm there will be 2 data analysis, the first using data with a minimum number of 1 item and transaction data with a minimum number of 2 items. The a priori algorithm analysis process is divided into 2 stages: high frequency pattern analysis and establishment of association rules with minimum support of 5% and confidence of 60%.
In October the transaction per day at Berkah Mart is 400 transactions or about 12,000 transactions in a month. The number of items reached 8,307 items, while the goods out per day is ¼ of items of goods that is 2,406. The item that was out in October was 74,586 items.

4.2. Data collection

Table 2 shown lists Support 1, itemset with minimum number of support is 10% can be shown in table 2 below:

Table 2. List of support values 1 itemset with minimum item count 1.

| Amount of goods                          | Amount | Support  |
|------------------------------------------|--------|----------|
| BaagusKrmrKentalMns 505 Gr               | 167    | 1.39%    |
| Fortune cooking oil 2 Ltr                | 247    | 2.05%    |
| Frisian Flag Cool-Choco 190 Ml           | 271    | 2.25%    |
| Indo Mie Kari Ayam 69 Gr                 | 511    | 4.25%    |
| Indomie Soto Medan 63 Gr                 | 419    | 3.49%    |
| Roma Kelapa Biscuit 300 Gr               | 178    | 1.48%    |
| Sofia Cooking Oil 2 L                    | 1562   | 13.02%   |
| Chocolatos 12Gr                           | 210    | 1.75%    |
| Small Chicken Egg 1.500                  | 544    | 4.53%    |
| Cow Sosis Champ                           | 339    | 2.82%    |
| Yellow Sugar 1Kg                          | 1292   | 10.76%   |
| Avena Cooking Oil 1 L                    | 467    | 3.89%    |

Furthermore, in Table 3, the value of support that reaches a minimum by 10% support for candidate combinations of 2 itemset can be shown in table 2 below:

Table 3. list of support value 1 itemset of combination 2 candidate itemset.

| Item of goods                  | Amount | Support  |
|-------------------------------|--------|----------|
| Oil Palm Sofia 2 L            | 1562   | 13.02%   |
| Yellow Sugar 1 Kg             | 1292   | 10.76%   |

While for support 2 itemset combination with minimum amount of support is 10%. Support on the combination of 1Kg Yellow Sugar (1.292 Transaction) and 2 L Sofia Cooking Oil (1,562 Transactions), with the sum of both combinations is 625 transactions. Based on the results of high frequency pattern analysis Support to seven combinations is shown in the following Table 4.
Table 4. List of Support Value 1 Itemset Combination Candidate Itemsets 2.

| No. | Type of goods       | Oct   | Nov   | Des   | Oct-Nov | Oct-Des | Nov-Des | All   |
|-----|---------------------|-------|-------|-------|---------|---------|---------|-------|
| 1   | Sofia Cooking Oil   | 13.02%| 13.37%| 14.22%| 12.71%  | 12.75   | 13.36   | 12.23 |
| 2   | Small Chicken Egg 1.500 | 10.31%| 11.23%| 11.32%| 10.61%  | 12.11   | 10.91   | 12.08 |
| 3   | Yellow Sugar 1 Kg   | 10.76%| 10.84%| 11.61%| 12.27%  | 11.32   | 11.38   | 9.97  |
| 4   | Choki-choki 1.000   | 9.83% | 09.33%| 10.72%| 10.03%  | 10.55   | 12.41   | 9.86  |

4.3. Establishment of association rules

Association rules that meet minimum confidence requirements by calculating the confidence of each item combination. Iteration stops when all items have been counted until there are no more item combinations. The confidence value of the combination of 1Kg Yellow Sugar and 2 L Sofia Cooking Oil is shown in Table 5 below:

Table 5. List of support value 1 itemset of combination 2 candidate itemset.

| Combination of 2 Items                                             | Support | Confidence |
|--------------------------------------------------------------------|---------|------------|
| IF buy Sofia Cooking Oil 2L, THEN will buy yellow sugar about 1Kg  | 13.02%  | 48.37%     |
| IF buy yellow sugar about 1Kg, THEN will buy sofia cooking oil     | 10.76%  | 40.01%     |

4.4. Testing and validation

This testing is done to know the truth of result of calculation and analysis using application of Tanagra 1.4 and Weka 3.7.4 by using transaction data of goods with at least 1 item and transaction data of goods with minimum 2 items with a minimum support 10% and minimum confidence 40%. The smaller the minimum value of support and minimum confidence that input will be more rules to be generated and vice versa the minimum value of support and minimum confidence that input will be less rule produced. Similarly, based on the analysis there are some things that need to be done arrangement is always close together that is:

Table 6. List of support value 1 itemset combination candidate itemsets 2.

| No. | Type of goods       | Oct   | Nov   | Des   | Oct-Nov | Oct-Des | Nov-Des | All   |
|-----|---------------------|-------|-------|-------|---------|---------|---------|-------|
| 1   | Sofia Cooking Oil   | 13.02%| 13.37%| 14.22%| 12.71%  | 12.75   | 13.36   | 12.23 |
| 2   | Small Chicken Egg 1.500 | 10.31%| 11.23%| 11.32%| 10.61%  | 12.11   | 10.91   | 12.08 |
| 3   | Yellow Sugar 1 Kg   | 10.76%| 10.84%| 11.61%| 12.27%  | 11.32   | 11.38   | 9.97  |
| 4   | Choki-choki 1.000   | 9.83% | 09.33%| 10.72%| 10.03%  | 10.55   | 12.41   | 9.86  |

4.5. Data Processing, Rule and Time Efficiency on FP-Growth

Like another Apriori of every experiment in the determination of rule, it is obtained the maximum results which have a resemblance between a priori rule and FP-Growth. The FP-Growth algorithm produces the same exact with a priori rule in each test of each support and the confidence value with 2 itemset combinations (such as table 4) and on any combination experiment (such as table 5). In FP-Growth, it has different value only on decimal value such as If you buy 2V sofia cooking oil, it will buy has support value 13.11% and Confidence 48.42%. And the rule If buying, it will buy sofia cooking oil 2L has a value of support 10.81% and Confidence 40.13%. Compared to Apriori, FP-
Growth has a higher support and confidence values than each rule, but in a theory, there is no significant difference if the difference between two algorithms is less than 1.

Regardless of the rule generated by the Apriori and FP-Growth, this study also compares the time efficiency required to build a second rule algorithm. In the experiment, resulted the time difference between the two algorithms based on 7 experiments, such as table 6 below

| Experiment | Dataset (Month) | Apriori (Secs) | FP-Growth (secs) |
|------------|-----------------|----------------|------------------|
| 1          | Oct             | 211.4          | 78.3             |
| 2          | Nov             | 262.6          | 123.5            |
| 3          | Des             | 209.3          | 69.4             |
| 4          | Oct – Nov       | 285.5          | 126.6            |
| 5          | Oct – Des       | 277.8          | 80.5             |
| 6          | Nov – Des       | 292.0          | 83.6             |
| 7          | Oct – Nov – Des | 330.3          | 131.7            |

The minimum time comparison used to form the rule of each algorithm which has a very long-time difference. It is recorded in seconds; the FP-Growth algorithm takes only 25% of the time which it takes to process 1 iteration in a rule, while Apriori uses 75% minimum time for 1 rule formation process.

Therefore, as research done by some researchers [17] and [18] suggests that FP-Growth is more optimal in case of market base analysis in terms of process speed. This research [19] states that the most effective association lure algorithm is FP-Growth. So, in this study that use 7 experiments also produce similar to previous research.

5. Conclusion
The proposed FP-Growth algorithm is more effective in generating the frequent and informative association rules to find out the consumer spending patterns at Berkah Mart in Pekanbaru compared to the Apriori Algorithm. As for the customer satisfaction, the related products are placed together to give the customers a logical view to select items they might buy. Moreover, from the rules of the
customer association, the goods can be segmented separately to meet the specific needs of customers with cost effective by using some special promotions for the general group. The experimental results show that the application of FP-Growth and Apriori algorithm for the analysis of consumer spending pattern at Berkah Mart can increase the overall income or profit, but it is recommended to use FP-Growth algorithm which maximally have the process speed in rule form and have superior support and confidence value a priori algorithm.

Acknowledgment
This research was supported fully by the grant from Research Institute of Puzzle Research Data Technology Faculty of Science and Technology UIN Sultan Syarif Kasim Riau. A subsubsection. The paragraph text follows on from the subsubsection heading but should not be in italic.

References
[1] Elsa, “Perkembangan Perusahaan pada Bidang Bisnis Ritel,” J. Manag., 2014.
[2] A. Mishra and D. Mishra, “Customer relationship management: Implementation process perspective,” Acta Polytech. Hungarica, vol. 6, no. 4, pp. 83–99, 2009.
[3] D. Banerjee and D. S. Shivani, “Analysis of Literature Review of Consumption Pattern – An Important Indicator of Economic Development,” Int. J., vol. 1, no. 2, 2011.
[4] M. Alkadri, “Analisa Data Mining Menggunakan Market Basket Analysis untuk Mengetahui Pola Beli Konsumen,” J. SATIN - Sains dan Teknol. Inf., vol. 1 No. 2, pp. 32–40, 2015.
[5] D. B. M. Pekanbaru, “Laporan Tahunan Penjualan dan Transaksi Berkah Mart Periode 2017,” 2017.
[6] M. Berry and G. Linoff, Data Mining Techniques: for marketing, sales and customer relationship management. 2004.
[7] P. Pravallika and K. Narendra, “Analysis on Medical Data sets using Apriori Algorithm Based on Association Rules,” IJSRSET, vol. 4, no. 1, pp. 717–722, 2018.
[8] R. Gangurde, D. B. Kumar, and D. S. D. Gore, “Building Prediction Model using Market Basket Analysis,” Int. J. Innov. Res. Comput. Commun. Eng., vol. 5, no. 2, pp. 1302–1309, 2017.
[9] R. C. Dorf and R. H. Bishop, Modern control systems. Pearson. Pearson Education, Inc., 2011.
[10] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” Procedia Comput. Sci., vol. 85, no. Cms, pp. 78–85, 2016.
[11] F. Alfiah, B. W. Pandhito, A. T. Sunarni, D. Muharam, and P. R. Matusin, “Data Mining Systems to Determine Sales Trends and Quantity Forecast Using Association Rule and CRISP-DM Method,” Int. J. Eng. Tech., vol. 4, no. 1, pp. 186–192, 2018.
[12] M. Madaio, R. Lasko, J. Cassell, and A. Ogan, “Using Temporal Association Rule Mining to Predict Dyadic Rapport in Peer Tutoring,” Int. Conf. Educ. Data Min., pp. 318–323, 2017.
[13] L. Khikmah, H. Wijayanti, and U. D. Syafitri, “Modeling Governance KB with CATPCA to Overcome Multicollinearity in the Logistic Regression Modeling Governance KB with CATPCA to Overcome Multicollinearity in the Logistic Regression,” J. Phys. Conf. Ser., 2017.
[14] R. Kanapaka, “Association Rule Mining using Apriori algorithm For food dataset,” Int. J. Comput. Appl., vol. 112, no. 4, p. 8887, 2015.
[15] D. Hunyadi, “Performance comparison of Apriori and FP-Growth algorithms in generating association rules,” Proc. Eur. Comput. Conf., pp. 376–381, 2011.
[16] M. Kavitha and M. S. T. T. Selvi, “Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons,” Int. J. Comput. Sci. Trends Technol., vol. 4, no. 4, pp. 161–164, 2016.
[17] D. A. I. J. I. Afzal, D. A. I. Umar, and D. K. Ullah, “Analyzing Apriori and FP-Growth Algorithm on an Arabic Corpus,” Int. J. Adv. Res. Comput. Sci. Softw, vol. 6, no. 1, pp. 147–150, 2016.
[18] C. Narula and G. Sikka, “Comparison of Apriori and Parallel FP Growth over Single-node and Multi-node Hadoop Cluster,” Proc. World Congr. Eng. Comput. Sci., vol. I, 2016.
[19] M. S. Mythili and A. R. Mohamed Shanavas, “Performance Evaluation of Apriori and FP-Growth Algorithms,” *Int. J. Comput. Appl.*, vol. 79, no. 10, pp. 34–37, 2013.