Pattern of E-marketplace Customer Shopping Behavior using Improved Tabu Search and FP-Growth Algorithm

Ayu Meida[1], Dian Palupi Rini[2], Sukemi[3]
[1,2,3]Faculty of Computer Science, Universitas Sriwijaya, Jl Srijaya Negara, Palembang, 30139, Indonesia

Article Info

ABSTRACT

Pattern of customer shopping behavior can be known by analyzing market cart. This analysis is performed using Association Rule Mining (ARM) method in order to improve cross-sale. The weakness of ARM is if processed data is big data, it takes more time to process the data. To optimize the ARM, we perform merging algorithm with Improved Tabu Search (TS). The application of Improved TS algorithm as optimization algorithm for preprocessing datasets, data filtering, and sorting data closely related products on sales data can optimize the ARM processing. The method of Association Rule Mining (FP-Growth) to determine frequent K-itemset, Support value and Confidence value of data which is already sorted on TS is based on patterns which often appear in the dataset so it generates rules as reference of decision making for company. To measure the level of power of rule which has been formed, the Lift Ratio value was calculated. Based on the calculation of 97 rules produced, the lift ratio produces values > 1 of 82.54% and based on processing time, it produces the fastest data search in 1.66 seconds. When compared with previous research that uses the hybrid method, for data retrieval based on processing time, it produces the fastest data search within 12.3406 seconds, 150 seconds and 50 seconds. Previous studies have only compared the processing time of data searching without regard to validation / accuracy of data search. The test results in this study obtained more optimal results than when compared with the results of previous studies, namely in time efficiency and data mining in real time and more accurate data validation. As a conclusion, the resulting rule can be used as a reference in understanding shopping behavior patterns customer on the E-Marketplace.

Corresponding Author:

Dian Palupi Rini,
Faculty of Computer Science,
Universitas Sriwijaya,
Jl. Srijaya Negara, Palembang 30139, Indonesia
Email: dprini@unsri.ac.id

1. INTRODUCTION

Customer is core element of any business model. Customer satisfaction is becoming the most important factor to support the success of a company. Every customer has different scheme in order to meet their necessity, which will affect their decision when buying a product. However, a group of customers still have some similarities of shopping behavior pattern in consuming product to maximize their satisfaction [1, 2]

The correlation between product's type that purchased by many customers can be analyzed to configure the pattern of customer shopping behavior [3].

In marketing area, customer shopping behavior pattern can be known using market cart analysis method [4]. Association rule mining is one of popular market cart analysis method that used customer shopping behavior. The problem of behaviors can be minimized, so it can improve cross-sales [5].

Association Rule Mining is the most important association's technique in data mining [6, 7]. In the transaction database can be found pattern, correlation, causal or association which occur between set of items or product by using Association Rule Mining [8] [9].

Application of Association Rule Mining on market cart analysis was tested on several research in various fields [10], for examples:

1. Research on Web Data using Swarm Optimization algorithm and Tabu Search method using Association Rule Mining. This result showed that the application of this method is exceed in time efficiency and mining data in real time [11]

2. Research on Supermarket using Association Rule Mining method, with main objective to analyze market cart. ARM searching for information in database on sale transaction with goal to compare the performance, so this research can be beneficial to business intelligence. [6].

3. Research on Online Music company applying Association Rule Mining method on their research, with goal to provide recommendation accurately and efficiently to the user based on user item mapping of activities and related tags with the item and determine similar preference in the same category. Results of the study prove that this method has more optimal performance compared to other methods [12].

Association Rule Mining method is most widely used as data mining method which used for analyzing market cart on transaction data in database with different case studies. Several studies have combined association rule mining method with several metaheuristic method and the results are quite good in accuracy compare with the standard method [11, 13]. Based on the previous research, Association Rule Mining method has combined with Taboo Search Algorithm method. Application of these hybrid method are better in improves time efficiency and search data in real time [13]. The previous combine methods have not found yet in E-Marketplace case. So, based on the success of hybrid ARM and Taboo search, the proposed research is applying the hybrid method on E-Marketplace case study by processing sales data in the database.

Association Rule Mining (ARM), one of the most commonly used is Frequent Pattern Growth (FP-Growth) algorithm process generate rules which are faster and more accurate [14]. ARM will optimize the shopping cart data on E-Marketplace using FP-Growth algorithm to search more detailed information based on variant of purchased product and understanding customer shopping behavior. This is aim to additional value as promotion for customer by recommending purchase history on shopping cart customer data, as well as marketing strategy to supply product which are most attractive to the customer. Thus, marketing can improve service to satisfy the customer in their shopping activity.

2.   RESEARCH METHOD
2.1. Data Mining
Data mining is one of the techniques for data mining or searching hidden information in database, some of the main function of data mining are association, prediction, clustering, estimation and classification [7, 15-17] .
2.2. Association Rule Mining
Association Rule Mining is one of the market cart analysis method which have strong relationship with study database on customer transaction data to figure out pattern of the customer shopping behavior, so it can be known that the products purchased at the same time [18]. Value of Support (supporting value) and Value of Confidence (certainty value) are parameters of Association Rule. Support value is percentage value of combined product item in the database, while confidence value is value to determine the strength of relation between product item on Association Rule [1, 2].

FP Growth Algorithm used for search Support value and Confidence value of sorted data based on pattern which often appear in the dataset, so it can be known which items that might come up (Support) or items which are exactly appear (Confidence) when data has been related. Pseudocode from FP Growth is explained as follows:

General Algorithm of *FP-Growth* :
1. *FP-Growth* (Tree, α)
2. **For** every (a dataset in margin above tree)
3. **Do** { β := a dataset **U** α }
4. Rising (β *with support* = dataset.support) construct condition of β against basic pattern and condition β against βtree FP-Tree)
5. **IF** βtree ≠ Ø
6. **Then** called *FP-growth* (βtree, β)

Calculation of support values can be seen in equations 1 and 2 [19] as follows:
To obtain the value of support for 1 item, use the following formula:

$$Support \ (A) \quad = \frac{\sum Transaction \ Contain \ A}{\sum Transaction} \qquad\qquad (1)$$

To obtain support values from 2 items use the following formula:

$$Support \ (A, B) = \frac{\sum Transaction \ Contain \ A \ and \ B}{\sum Transaction} \qquad\qquad (2)$$

Calculation of confidence values can be seen in the equation 3 [17] the following :

$$Confidence \quad = P\,(\,B\mid A\,) = \frac{\sum Transaction\ Contain\ A\ and\ B}{\sum Transaction\ Contain\ A} \qquad (3)$$

2.3. Improved Tabu Search

Algorithm Improved Tabu Search will be added to the algorithm for preprocessing dataset, data filtering, and sorting data related product on sales data so that the dataset which still empty or there is field which is not useful can be cleaned, like the Pseudocode in point 8:

1. S ← some of Solution initial candidate.
   a better ← S.
2. L ← { } List of Tabu length Maximum 1
3. I ← 1.
4. **While** I < Max-Iter and not stop **do**
5. Enqueue S to L
6. S ← Best Neighbor (S).
7. Alter best if quality(best) < Quality (S)
8. Check *Row*.
       If Null, delete.
       While row <> Null
       Filter data
       Sort data
9. End while

2.4. Dataset

This research uses dataset Online Retail from repository data bank University of California (UCI Repository Machine Learning Benchmark). The dataset will be pre-processed to eliminate the data empty (missing value), noisy data or unreasonable value. The attributes of the dataset that used in this research are InvoiceNo, InvoiceDate, CustomerID, Description, StockCode, Quantity. The amount of data used is 1000 data.

2.5. Framework

Design and implementation on this research have several steps which will be perform, in order to obtain the expected results. The framework of this research is shown in figure 1.
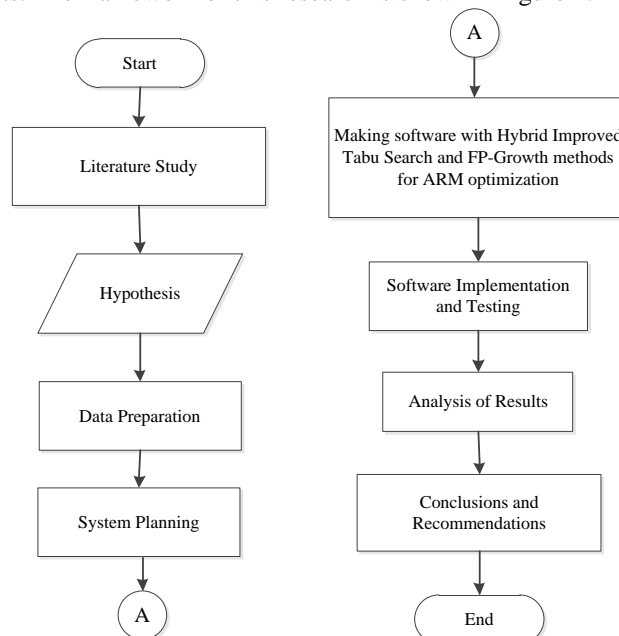

Figure 1. Flowchart of Research Framework

2.6. Combination of Improved TS Algorithm and ARM (FP-Growth)

Improved TS Algorithm is used for pre-processing datasets, data filtering, and sorting data that have closely related product on sales data. FP-Growth algorithm is used to process ARM, to obtain frequent k-itemset, Support value and Confidence value of data which is already sorted by Improved Tabu Search

algorithm, so it can be known which items that might come up (Support) or items that certainly appear (Confidence) when those data has been related and generate rules as final result of processing data which are useful for taking decisions of company. ARM (FP-Growth) Calculation is started by determining minimum Support value and Confidence value of sales data which taken from database. After Support value and Confidence values inputted and processed, then the system will determine K-Itemset which can be formed. The result of K-Itemset which has been processed by the system, will be stored as rules of the Association. So that, it can be a reference in determine stock product which will be provided to discover the pattern of necessity of the customer. Figure 2 shows a flowchart of combination two algorithm which used in this research.
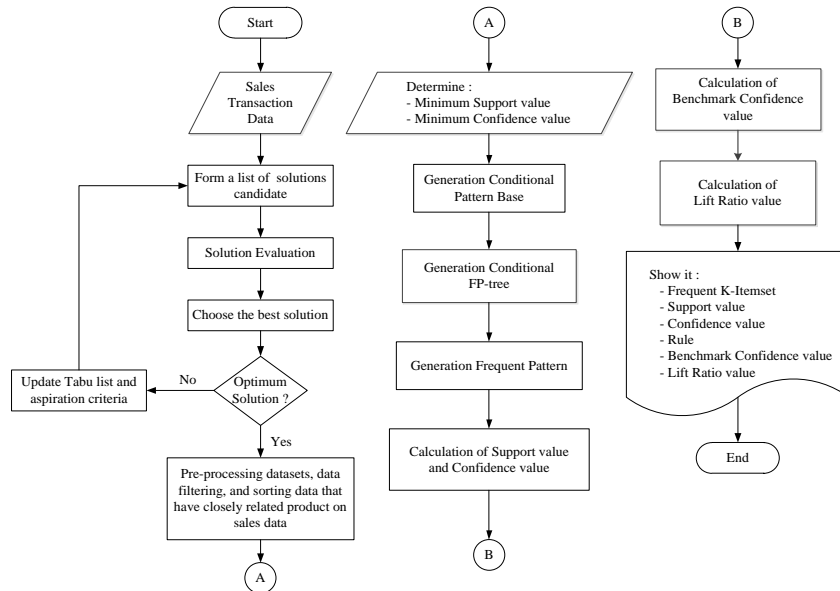

Figure 2. *Flowchart of* combination TS-ARM (FP-*Growth*)

The combination of the two algorithms are used in accordance with its functions respectively. So that it can be known which items that might come up (Support) or items that certainly appear (Confidence) when the data has been related and generate rules as a reference for decision making.

3. RESULTS AND DISCUSSION
3.1. Discussion

The system of customer shopping behavior pattern has been developed. the system is built in visual form. Each interface has its own function in its task to build consumer shopping behavior patterns using Improved TS and FP-Growth algorithms. Datasets which are uploaded must be in format xlsx. After dataset upload it will be processed with Improved TS algorithm. The number of original datasets is 1000 data, after being processed into 976 data.

a. Determine Frequent Itemset

After 5% minimum support value and 10% minimum confidence processed, then the frequent itemset with 50 product data entry will be shown as illustrated in Figure 3.

No	InvoiceNo	StockCode	Description	Count
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	10
2	536365	71053	WHITE METAL LANTERN	5
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	5
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	9
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	6
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	5
8	536366	22633	HAND WARMER UNION JACK	10
9	536366	22632	HAND WARMER RED POLKA DOT	11
10	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	7

Showing 1 to 10 of 50 entries   Previous   1   2   3   4   5   Next

Figure 3. Frequent Itemset sample

b. Determine Association Rule

After the minimum support value and minimum confidence value has been processed and shows the frequent item set, then Association rules will be generated on menu Association Rule. Association Rules will display Rule, Support value, Confidence value, Confidence Benchmark Value, Lift Ratio value and speed of Data Load Times. The results of the experiment use 5% minimum support value and 10% minimum confidence which generate 50 Association rule of sample data. the period for the process is 6.45 seconds. The results of association are shown in Figure 4.

No	Kode Rule	InvoiceNo	StockCode	Description	Count	Minimal Support	Minimal Confidence	Association	Nilai Support	Nilai Confidence	Confidence Benchmark	LiftRatio	Data Load Times
1	1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	10	0.05	0.1	Jika membeli WHITE HANGING HEART T-LIGHT HOLDER maka akan membeli WHITE METAL LANTERN	3.08	0.5	0.02	1.25	6.45 seconds
2	2	536365	71053	WHITE METAL LANTERN	5	0.05	0.1	Jika membeli WHITE METAL LANTERN maka akan membeli CREAM CUPID HEARTS COAT HANGER	3.08	0.5	0.02	1.25	6.45 seconds

Showing 1 to 10 of 50 entries   Previous 1 2 3 4 5 Next

Figure 4. Association Rules sample

3.2. Testing

On this research will perform testing using Benchmark Confidence and Lift Ratio. Lift Ratio assess the strength level of generated rule from calculation support value and confidence value. Lift Ratio value is used to determine generated rule which is valid or invalid [19, 20]. Lift Ratio can be calculated using formula below 5 and 6 [15]:

$$Lift\ Ratio = \frac{Confidence}{Benchmark\ Confidence} \tag{5}$$

The Benchmark of Confidence value can be calculated using formula below:

$$Benchmark\ Confidence = \frac{\sum Transaction\ Item\ in\ Consequent}{\sum Transaction\ in\ Database} \tag{6}$$

The calculation result of support value is obtained as follow:
1. If minimum Support value = 5% and minimum Confidence value = 10% then generated 50 rules
2. If minimum Support value = 7% and minimum Confidence value = 10% then generated 25 rules
3. If minimum Support value = 10% and minimum Confidence value = 15% then generated 10 rules
4. If minimum Support value = 12% and minimum Confidence value = 15% then generated 5 rules
5. If minimum Support value = 15% and minimum Confidence value = 20% then generated 3 rules
6. If minimum Support value = 15% and minimum Confidence value = 25% then generated 3 rules
7. If minimum Support value = 16% and minimum Confidence value = 25% then generated 1 rules
8.

Based on the testing above, if minimum support value above 16 % then it will not generate any rule. The Calculation of lift ratio which generate 97 rules above is presented as follows:

Table 1. Result of Lift Ratio Testing

Lift Ratio	Value	Percentage
1	17	17,5%
1,25	29	29,9%
1,5	51	52,6%

The higher the minimum support and confidence values processed, the fewer frequent itemset and rule are formed. All rules produced in this study have a lift ratio value of 17.5%, a lift value of 1.25 of 29.9% and a lift ratio of 1.5 of 52.6%. A rule is declared valid if the lift ratio value is not lower than 1 and the strength of the rule will be more valid if it is > 1, so it can be concluded that the rule in the above experiment is very valid at 82.5%.

Based on the test results above, with a sample dataset of 1000 data using a hybrid improved Tabu Search method and FP-Growth for optimization of the Rule Mining Association with an E-Marketplace case study based on processing time, producing the fastest data search within 1.66 seconds and validation testing the data using lift ratio reaches a percentage of up to 82.5%. Previous studies using hybrid methods [9, 11, [11] for data search based on processing time, resulting in the fastest data search within 12.3406 seconds, 150 seconds and 50 seconds. Previous studies have only compared the processing time of data searching without regard to validation / accuracy of data search. So it can be concluded that by using the Hybrid Improved Tabu Search method and FP-Growth for optimization of the Rule Mining Association in the E-Marketplace case study, more optimal results are obtained compared to previous method research, namely in time efficiency and data mining in real time and data validation more accurate.

4. CONCLUSION

Based on the results in this research, it can be concluded that Tabu Search algorithm is used for filtering and sorting related product on sales data which are expected can optimized ARM (FP-Growth) process, Tabu Search and ARM (FP-Growth) were quite good in terms of association and speed of data processing, Shopping cart data analysis using ARM (FP-Growth) can generate rules which can help increase sale, manage stock inventory of product, and can arrange sales promotion strategy in online sales, The higher number of minimum support and minimum confidence which is used then the smaller number of frequent itemset and rule which is formed, All rule that resulted in this research have lift ratio > 1 of 82.5% and producing the fastest data search within 1.66 seconds then the rule can be used as reference in understanding customer shopping behavior pattern on the E-Marketplace.

REFERENCES

[1] F. Kurniawan, B. Umayah, J. Hammad, S. M. S. Nugroho, and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviours by Way of Transaction Data," *Knowledge Engineering and Data Science,* vol. 1, no. 1, pp. 20-25, 2017.

[2] A. K. Prasidya and C. Fibriani, "Analisis Kaidah Asosiasi Antar Item Dalam Transaksi Pembelian Menggunakan Data Mining Dengan Algoritma Apriori (Studi Kasus: Minimarket Gun Bandungan, Jawa Tengah)," *JUTI: Jurnal Ilmiah Teknologi Informasi,* vol. 15, no. 2, pp. 173-184, 2017.

[3] A. Griva, C. Bardaki, K. Pramatari, and D. Papakiriakopoulos, "Retail business analytics: Customer visit segmentation using market basket data," *Expert Systems with Applications,* vol. 100, pp. 1-16, 2018.

[4] M. Kaur and S. Kang, "Market Basket Analysis: Identify the changing trends of market data using association rule mining," *Procedia computer science,* vol. 85, pp. 78-85, 2016.

[5] J. Phani Prasad and M. Mourya, "A Study on Market Basket Analysis using a Data Mining Algorithm," 2013.

[6] S. Abdulsalam, K. Adewole, A. Akintola, and M. Hambali, "Data Mining in Market Basket Transaction: An Association Rule Mining Approach," *International Journal of Applied Information Systems (IJAIS),* vol. 7, no. 10, pp. 15-20, 2014.

[7] P. Rajendran and M. Madheswaran, "Hybrid medical image classification using association rule mining with decision tree algorithm," *arXiv preprint arXiv:1001.3503,* 2010.

[8] C. Kaur, "Association rule mining using apriori algorithm: a survey," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* vol. 2, no. 6, 2013.

[9] M. Abdel-Basset, M. Mohamed, F. Smarandache, and V. Chang, "Neutrosophic association rule mining algorithm for big data analysis," *Symmetry,* vol. 10, no. 4, p. 106, 2018.

[10] T. A. Kumbhare and S. V. Chobe, "An overview of association rule mining algorithms," *International Journal of Computer Science and Information Technologies,* vol. 5, no. 1, pp. 927-930, 2014.

[11] Y. Djenouri, H. Drias, and A. Chemchem, "A hybrid bees swarm optimization and tabu search algorithm for association rule mining," in *2013 World Congress on Nature and Biologically Inspired Computing*, 2013, pp. 120-125: IEEE.

[12] M. K. Najafabadi, M. N. r. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," *Computers in Human Behavior,* vol. 67, pp. 113-128, 2017.

[13] S. Kaur and M. Goyal, "Fast and robust hybrid particle swarm optimization tabu search association rule mining (HPSO-ARM) algorithm for web data association rule mining (WDARM)," *J. Adv. Res. Comput. Sci. Manag. Stud,* vol. 2, pp. 448-451, 2014.

[14] G. Gunadi and D. I. Sensuse, "Penerapan metode data mining market basket analysis terhadap data penjualan produk buku dengan menggunakan algoritma apriori dan frequent pattern growth (fp-growth): studi kasus percetakan pt. Gramedia," *Telematika MKOM,* vol. 4, no. 1, pp. 118-132, 2016.

[15] M. Brilliant and D. Handoko, "Implementation of Data Mining Using Association Rules for Transactional Data Analysis," in *Prosiding International conference on Information Technology and Business (ICITB)*, 2018, pp. 177-180.

[16] M. Heydari and A. Yousefli, "A new optimization model for market basket analysis with allocation considerations: A genetic algorithm solution approach," *Management & Marketing,* vol. 12, no. 1, pp. 1-11, 2017.

[17] R. Sarno, R. D. Dewandono, T. Ahmad, M. F. Naufal, and F. Sinaga, "Hybrid Association Rule Learning and Process Mining for Fraud Detection," *IAENG International Journal of Computer Science,* vol. 42, no. 2, 2015.

[18] M. A. Valle, G. A. Ruz, and R. Morrás, "Market basket analysis: Complementing association rules with minimum spanning trees," *Expert Systems with Applications,* vol. 97, pp. 146-162, 2018.
[19] D. Despitaria, H. Sujaini, and T. Tursina, "Analisis Asosiasi pada Transaksi Obat Menggunakan Data Mining dengan Algoritma A Priori," *Jurnal Sistem dan Teknologi Informasi (JUSTIN),* vol. 4, no. 2, pp. 277-282, 2016.

BIOGRAPHY OF AUTHORS

Ayu Meida
Student of Magister Teknik Informatika Sriwijaya University. She finished her diploma of Manajemen Informatics in Politeknik Sriwijaya Palembang and under graduated of Information System in Sriwijaya University.

Dian Palupi Rini
Lecture of Magister Teknik Informatika Sriwijaya University. She finished her under graduated of Math in Sriwijaya University (2000), Magister of Computer Science in Gadjah Mada University (2003) and Doctor of Computer Science in Universiti Teknologi Malaysia (2017)

Sukemi
Lecture of Magister Teknik Informatika Sriwijaya University. He finished his under graduated of Electrical Engineering in Sriwijaya University (1992), Magister of Electrical Engineering in Institut Teknologi Bandung (1999) and Doctor of Electrical Engineering in Indonesian University (2016)