International Journal of Engineering & Technology, 7 (4.29) (2018) 115-121

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Recommendation System to Optimize Email Marketing Campaign Using Apriori Algorithm Case Study: Webeli.Com

Ratih Widayanti Kosaman[1], Djoko Harsono[2], Ibnu Hadi Sumitro[3]

[1,2,3]*Department of Information System STMIK Indonesia Jakarta Jakarta, Indonesia*

Abstract

Email marketing campaign has uniqueness in customer personalization, ability to raise the conversion rate of a website, and a high click-through rate. Unfortunately, email marketing campaign cannot always provide relevant contents for customers. To solve this problem, Webeli.com build recommendation system to find out what products are liked by the customers. The purpose of this research is using association rules technique to optimize email marketing campaign and to influence customers into purchasing particular products that relevant to their interest. The algorithm that used to analyze the association rules among products that purchased together by the customers in one shopping cart is apriori algorithm. This research produces a system that can provide product recommendations that relevant with customers' interest and can optimize email marketing campaign by increasing efficiency number of the email recipient, duration of delivery email, and cost of delivery email.

Keywords: *Email Marketing, Product Promotion, Recommendation Systems, Apriori Algorithm, Association Rules.*

1. Introduction

Webeli.com is an e-commerce startup that use email marketing campaign to promote merchants' products. Email marketing has more advantages than other online marketing tools, it has uniqueness in customer personalization, ability to raise the conversion rate of a website, and a high click-through rate. Unfortunately, email marketing cannot always provide the recipients with relevant contents [1].

To solve this problem, Webeli tries to build a system to provide product recommendation that relevant with customers' interest. The main goal of this research is using association rules technique to explicitly influence customers into purchasing particular products. Association rules technique can discover interesting associations between attributes contained in a database, all possible combinations of potentially interesting product grouping can be explored [2,3,4]. The algorithm used to build it is Apriori algorithm. Apriori Algorithm can be applied to support marketing strategy in a company. Information related to implementation promotion can be available quickly, so the management can make a decision with fast [5].

Pramudiya et al., in their research state that Apriori algorithm can be used as a recommendation tool for the products that will be purchased by the customers and it can be used to filter potential customers [6,7,8,9,10,11,12]. Sharif et al., also conduct a similar research the result is association rules can be used to send online ads about the latest products to the potential customers based on their profile [13]. In other research, Zulfikar et al. use apriori algorithm to increase the quality of promotion, the result is association rules that contain a combination of varied products as a reference for promoting the products [14]. This research has similarities to the three pieces of research that has discussed before. The differences are this research produces a system that can provide product recommendations that relevant with customers' interest and measures how much recommendation system can optimize email marketing campaign by increasing efficiency the number of email recipient, duration of delivery email, and cost of delivery email.

The first step, this research analyzes email marketing on the previous system by using a few of variables. The purpose of this analysis is to find out the problems that occur in email marketing activities on the previous system. Second, build a prototype system by applying apriori algorithm. Third, the prototype will be tested by input minimum support value and minimum confidence value. It produces product combination that can be recommended to the customers. The prototype will look for customers who have similarities of products that are in their purchase transactions history with a combination of products produced by the association rules. Then the system will send email marketing to the customers. Fifth, the prototype will be analyzed using six variables that are how to select products that will be sent through email marketing, how to select candidates who will receive email marketing, duration of email marketing delivery, email marketing delivery cost, and survey responses. The final step is comparing the analysis results of the previous system and the new system to find out how much efficiency of the new system.

2. Theory

2.1. Association Rules

Customer shopping patterns are very interesting to analyze. Some valuable information can be generated from it. The information can be used to support business activities such as product marketing, stock management, and customer relationship management. A methodology that can be used to analyze customers' shopping patterns is association analysis. It is a methodology used to find the association among products on the customers' purchase transaction data. For example, see Table 1.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table1: Purchase Transaction Data

TID (*T*)	Item (*I*)
1	Egg, Noodle
2	Egg, Crackers, Bread, Mineral Water, Noodle
3	Noodle, Crackers, Bread, Mineral Water, Jam
4	Egg, Noodle, Crackers, Bread
5	Egg, Noodle, Crackers, Mineral Water
6	Noodle, Crackers, Egg

In Table 1, noodles and eggs simultaneously exist in almost all transactions. It shows that noodles and eggs have strong association. There are some basic terminologies that should be known to determine association analysis such as [15]:

Binary representation

Purchase transaction data can be represented in a binary format using the value "0" and "1". Value "1" is given if the products exist in the purchase transaction data. Value "0" is given if the products do not exist in the purchase transaction data. Table 2 show binary representation from table 1.

Table2: Binary Representation From Table 1

TID	Mineral Water	Noodle	Bread	Jam	Crackers	Egg
1	0	1	0	0	0	1
2	1	1	1	0	1	1
3	1	1	1	1	1	0
4	0	1	1	0	1	1
5	1	1	0	0	1	1
6	0	1	0	0	1	1

Itemset and support count

There is variable *I*, and $I = \{i_1, i_2, i_3, ..., i_n\}$ defines all items in the transaction history data. There is variable *T*, and $T = \{t_1, t_2, t_3, ..., t_n\}$ defines all transactions contained in the database. Each transaction t_i contains items taken from *I* (eg: t_1 transaction contains an item i_2 and i_3). In association analysis, a set of empty item or a set that contains more than one item is called itemset (eg: {}, $\{i_2, i_3\}$, and $\{i_5\}$ called itemset). Empty set is an itemset that does not have any item. If an itemset contains *k* items, it is called *k*-itemset. Examples on transaction TID = 2, there is an itemset {*Eggs, Crackers, Bread, Mineral water, Noodles*}, then it can be called a 5-itemsets because it consists of five types of item).

Transaction width

Transaction width is defined as a number of items contained in a transaction. T_j transaction can be said containing an itemset X if X is a subset of t_j. For example, t_j transaction consists of itemset $\{i_1, i_2, i_3\}$, itemset $\{i_1, i_2\}$ is the subset of t_j, while the itemset $\{i_1, i_5\}$ is not the subset of t_j because i_5 is not included in t_j.

An important property from an itemset is support count. Support count is the number of transactions containing certain itemset.

Association rule

Association rule is an implication expression of $X \rightarrow Y$, where X and Y are disjoint itemsets, $X \cap Y = \emptyset$. The quality of association rules can be measured with support and confidence.

Support and Confidence

Support is how often a rule is applied on dataset, while confidence can determine the frequency of items in Y that appears in transactions containing X. Support is used to eliminate rules that do not interesting, while confidence is used to measure the reliability of conclusions made from the rules.

Association analysis is also often reffered as a market basket analysis. Basic methodology of association analysis is divided into two phases, namely:

Frequent itemset extraction

This phase is looking for a combination of items that meets the minimum requirements of support value. Support value of one item is obtained using the following formula:

$$Support\ (A) = \frac{number\ of\ transaction\ containing\ A}{total\ transaction}$$

While support value of two items is obtained using the following formula:

$$Support\ (A \cap B) = \frac{number\ of\ transaction\ containing\ A\ and\ B}{total\ transaction}$$

Association rule mining

After all frequent itemsets are found, the next step is finding an association rule. Association rule must meet the minimum support and minimum confidence. Confidence value of rules is derived from the following formula:

$$Confidence = P\ (B|A) = \frac{number\ of\ transaction\ cont}{number\ of\ transaction\ c}$$

2.2. Hypothesis

The hypotheses are as follows:

H1 = recommendation system can provide relevant recommendation products to the customers based on the similarity of purchased product and combination products in association rules.

H2 = recommendation system can optimize email marketing activities.

3. Methodology

3.1. Research Method

System development method used is prototyping method. The prototype will be built, tested, and then repeated as necessary until the prototype is accepted by Webeli. For the system design, the approach used is object-oriented approach.

3.2. Sample Method

Sampling technique used is purposive sampling. Purposive sampling is a technique based on specific criteria sampling. The num-

ber of sample is calculated using the slovin formula. Where n is the sample unit, *N* is the population, and *e* is the error value.

$$n = \frac{N}{1 + N(e)^2}$$

Samples are customers who have bought at least one transaction of which the purchase status is approved. If the number of Webeli's customers who have purchased at least one time and have status of transaction is approved is 182 people (*N*) with a sampling error rate of 5% (*e*), then the minimum number of samples (*n*) to be taken is 125 people.

3.3. Instruments

Data is collected by observation, interview, and questionnaire analysis. Instruments used are:
Instruments for observation method are operational business process and email marketing activities.
Instruments for interview method is the list of questions that will be given to the interviewees.
Instruments for questionnaire method is the list of questions that will be given to respondents. Likert scale range used is:

Table3: Likert Scala Range

Options	Rating
Very uninteresting	0 – 1,0
Not interesting	1,1 – 2,0
Neutral	2,1 – 3,0
Interesting	3,1 – 4,0
Very interesting	4,1 – 5,0

3.4. Testing Technique

Techniques that used to measure the relevance between the recommendation and the customer interest is the trend of rising percentage, using the formula (new data - old data) / (new data) * 100%. Old data taken from the result of questionnaires analysis from the previous system, while new data is taken from the result of questionnaire analysis from the new system. The test results will be presented on the comparison table.
Techniques that used to measure the efficiency of email marketing delivery is calculating the trend of rising percentage, using the formula (new data - old data) / (new data) * 100%. The efficiency testing is carried out by using the parameter of the number of email marketing recipients, the duration of email marketing delivery, and email marketing delivery cost. Old data is taken from the number of email marketing recipients, the duration of email marketing delivery, and email marketing delivery cost from the previous system, while the new data is taken from the number of email marketing recipients, the duration of email marketing delivery, and email marketing delivery cost from the new system.

4. Analysis

4.1. Previous System Analysis

There are six variables that used for analyze previous system, namely: 1) how to select products that will be sent through email marketing, 2) how to select candidates who will receive email marketing, 3) the number of email marketing recipients, 4) the duration of email marketing delivery, 5) the email marketing delivery cost, and 6) the results of survey responses toward customer product recommendation. The conclusion of previous system condition is shown in Table 4.

Table4: Conclusion of Previous System Testing

Variables	**Results**
How to select products	Random
How to select candidates	Random
The number of email recipients	All recipients in contact list. Total recipients on the contact list of "Webeli Members" are 926 people.
Duration of email delivery	Delivery for one sale promotion to the contact list of "Webeli Members" requires 27 minutes and 30 seconds.
Email marketing delivery cost	Delivery for one sale promotion to the contact list of "Webeli Members" costs US $ 0.0001 * 926 = US $ 0.0926, equivalent to Rp1.214,91, -.
Match between product recommendation and customers' interests	Value range of the match is 3 or equal to matched enough.
Variables	Results

After analyzing the operational business process and email marketing campaign activities, we can conclude that several problems occurred are:
There is no system that can provide relevant product recommendation to the customers' interests or wishes based on customers' purchase transaction history.
Email marketing is sent to all subscribers without filtering email marketing recipients. It can affect duration and cost of email delivery.
From the analysis, it is suggested that Webeli.com build a recommendation system. The recommendation system can be useful for Webeli.com in establishing product marketing strategy and making an email marketing activities more focused. Here is the proposed business process email marketing:
The content writer makes email marketing templates using HTML.
The content writer creates contents of email marketing product
The content writer provides contents of email marketing to the Administrator.
The administrator receives contents of email marketing.
The administrator runs the system on to the stages of frequent *k*-itemsets extraction, generates association rules, and transmits the association rules to the Interspire Email Marketer application.
The administrator runs the Interspire Email Marketer application to send email marketing.
The administrator sends email marketing to the customers in accordance with the recommendations given.
Customers receive email marketing from Webeli.com.

4.2. Functional Requirement Analysis

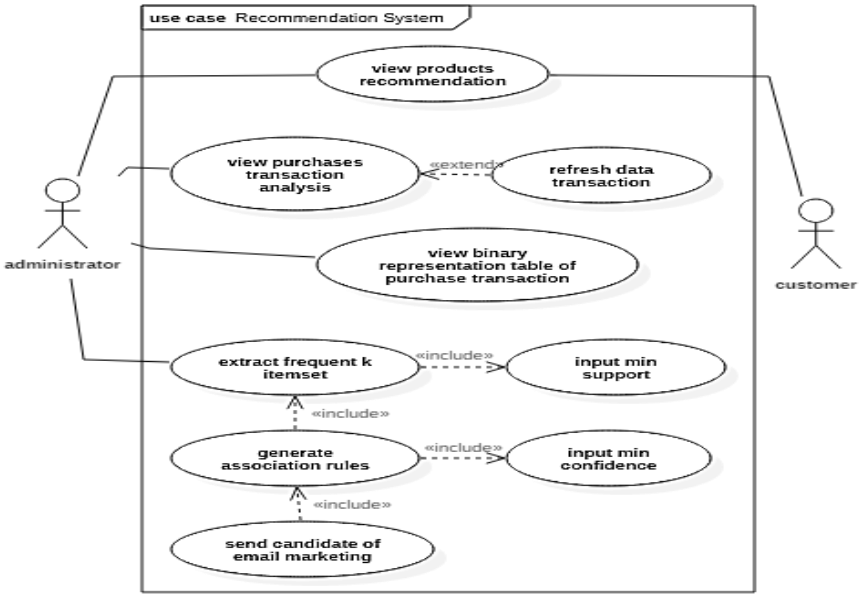
Functional requirements analysis is conducted to determine the functional requirement of the system. This requirement is derived from an interview with the Webeli.com. Table 5 shows the specific requirement of the built system software.

Table5: Specific Requirement of Software

Number	Specification
Requirement-01	System should display sales transaction analysis on a table.
Requirement-02	System should display the table of binary representation of sales transaction data.
Requirement -03	System should be able to search frequent k-itemset, to display frequent k-itemsets candidate, and to display the frequent k-itemset extraction.
Requirement -04	System should display the association rules generated from frequent k-itemset extraction.
Requirement -05	System should send the candidate of email marketing recipients and product names to Interspire. Customers who have ever bought one of the products listed on the association rules will be a strong candidat to receive the email marketing.
Requirement -06	System should display product recommendations on the product detail page.

4.3. Use Case Diagram

Use Case diagram is used to model the specific requirements of software shown in Table 5.

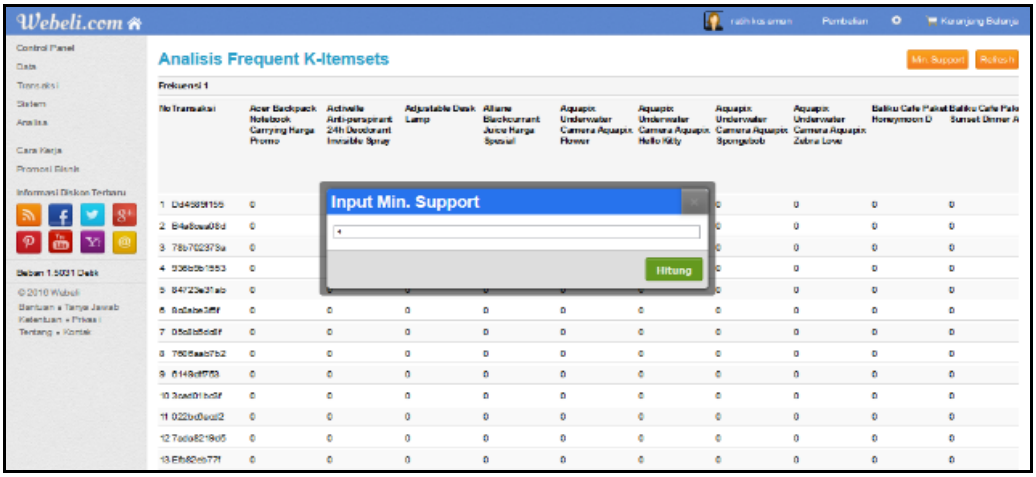


Picture1: Use Case Diagram

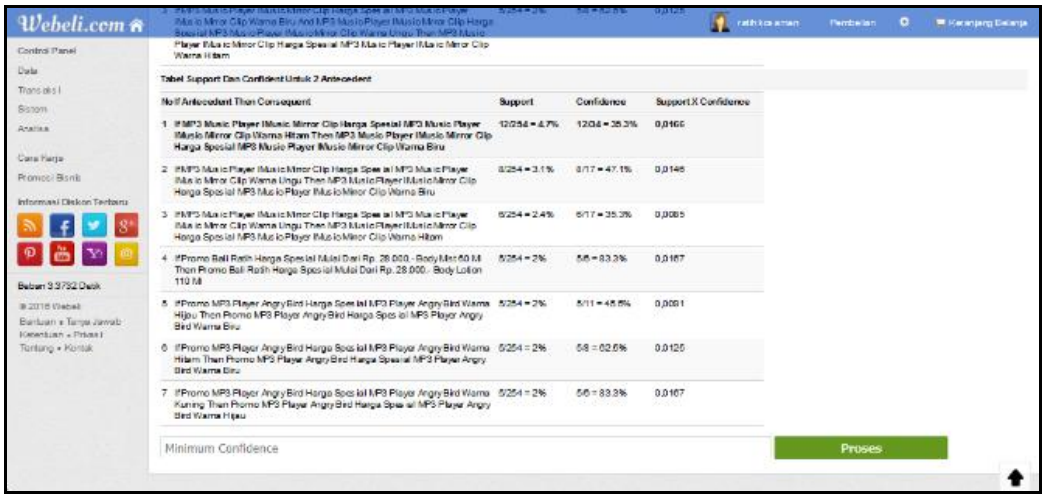
5. Results

5.1. Prototype Implementation

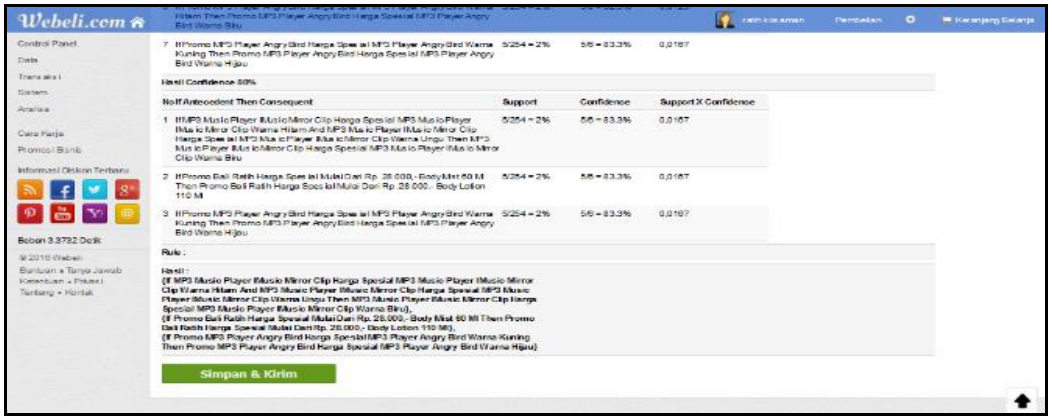
The following previews show the built of recommendation system.



Picture2: Input Minimum Support



Picture3: Input Minimum Confidence



Picture4: Association Rules Mining

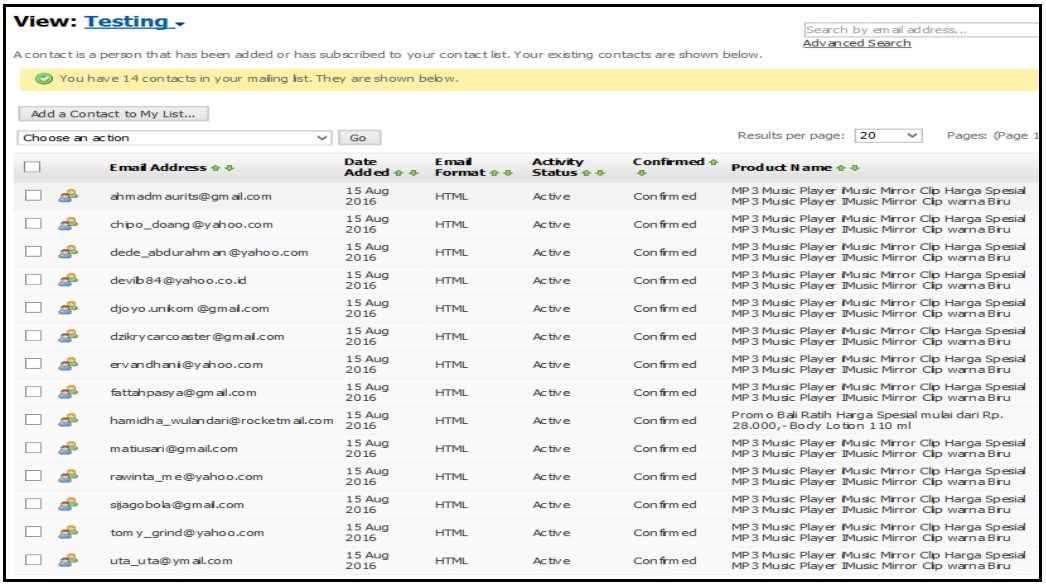
Using minimum support 40% and minimum confidence 80%, the association rules contain three combination of products that can be recommended to customers, namely:

If customers have purchased "MP3 Player Mirror Clip Black" and "MP3 Player Mirror Clip Purple" then the customers will be recommended to buy "MP3 Player Mirror Clip Blue".

If customers have purchased "Body Mist Bali Ratih 60 ml" then the customer will be recommended to buy "Body Lotion Bali Ratih 110 ml".

If customers have purchased "MP3 Player Angry Bird Yellow" then the customer will be recommended to buy "MP3 Player Angry Bird Green".

The system can find 14 customers who have similarity purchase transaction history with the three combination products above. The lists follow:



Picture5: Email Recipients in New System

5.2. Recommendation System Testing

There are six variables that used for testing previous system, namely: 1) how to select products that will be sent through email marketing, 2) how to select candidates who will receive email marketing, 3) the number of email marketing recipients, 4) the duration of email marketing delivery, 5) the email marketing delivery cost, and 6) the results of survey responses toward customer product recommendation. The conclusion of new system is shown in Table 6.

Table6: Recommendation System Testing

Variables	Results of Analysis
How to select products	The results from generating association rules. From 40% minimum value support and 80% of minimum confidence, it is gained 3 recommended products.
How to choose candidates	Based on the similarity between products that have been purchased and those listed on association rules.
The number of email recipients	There are 14 customers who have purchased the same type of products as those listed on the association rule.
Duration of email delivery	The duration for sending email marketing "iMusic Mirror MP3 Music Player Clip" is 25 seconds. Meanwhile, the duration for sending email marketing "Body Lotion 110 ml" is 7 seconds, so the total is 32 seconds.
Email marketing delivery cost	Sending email marketing "iMusic Mirror MP3 Music Player Clip" and "Body Lotion 60 ml Ratih Bali" to 14 people costs US $ 0.0001 * 14 = US $ 0.0014, which is equal to Rp18,36-.
Match between product recommendation and customers' interests	Based on one of the 13-customer's responses, the value range is 4 or equal to matched.

5.3. Comparison of Previous System and New System

In this section, we will discuss the comparison of previous system and new system. The purpose of this comparison is to prove the effectiveness of the new system by calculating the efficiency of how to select product, how to select email recipients, the number of email recipients, the duration of email delivery, the email delivery cost, and the match of product recommendations. This comparison is also used to prove the hypothesis of the study. Table 7 shows the comparison of the previous system and the new system.

Table7: Comparison of the Previous System and the New System (1)

Variable	Product Selection	Email Recipients Selection
Previous System	Random, example: Keripik Singkong Krispi Renyah	Random,example: Webeli Member List
New System	Based on association rules: MP3 Player Mirror Clip Blue, Body Lotion Bali Ratih 60 ml	Based on association rules with customer's purchases history: MP3 Player Mirror Clip Blue sent to 13 people, Body Lotion Bali Ratih 60 ml sent to 1 person
Efficiency	Product selection is more directed.	Filtering email recipents from all members in the list to only several members is carried out.

Table8: Comparison of the Previous System and the New System (2)

Variable	The Number of Email Recipients	Duration of Email Marketing Delivery
Previous System	926 customers	27 minutes and 30 seconds = 1650 seconds
New System	14 customers	MP3 Player Mirror Clip Blue needs 25 seconds and Body Lotion Bali Ratih 60 ml needs 7 seconds, total = 32 seconds
Efficiency	(previous data-new data)/new data* 100 = (926 – 14)/14 *100 = 65, 14 * 100 = 6514%	(previous data-new data)/new data * 100 = (1650 – 32)/32 *100= 50,56 * 100= 5056%

Table9: Comparison of Previous System and New System (3)

Variable	Email Marketing Delivery Cost	Match of Product Recommendation
Previous System	Rp1.214,91	Range value=3 or Neutral
New System	Rp18,36,-	Range Value=4 or Interesting
Efficiency	(previous data – new data)/new data* 100 = (1214,91-18,36)/18,36 *100= 65,17*100= 6517%	New system is one level higher than previous system

Based on the comparison shown on Table 7, 8, 9, it can be seen that the efficiency of the number of email recipients in the new system is 6514% better than the previous system. The efficiency rate of the number of email recipients gives an impact toward the duration of email delivery. The efficiency of the duration email delivery of the new system is 5056% better than the old system. The efficiency rate of the number of email recipients also gives an impact toward email delivery cost. The efficiency of delivery cost on the new system is 6517% better than the previous system. From this comparison, it can also be seen that the selection of products that will be recommended is more focused. Besides, the system has also successfully filtered the recipients based on the similarity between the products that have been purchased and the combination products in the association rules. In addition, from the comparison results, it is shown the increasing rate of the match between the product recommendation and the customers' interest. The rate of the match of product recommendation on the new system goes up one level from rating 3 (neutral) to 4 (interesting). There are obstacles in these surveys, that is, not all the customers fill out and respond to the survey.

6. Conclusion

Recommendation system can find 14 customers who has relevant purchase transaction history with the three products above. The relevance is based on similarity between the products that have ever been bought by the customers and the products listed on the association rules. In addition, the survey results that recommendation system can deliver products promotion based on customers'

interests. This result is proved by the increasing rate of the match of product recommendation and customers' interests from rating 3 to 4. There are obstacles in these surveys, that is, not all the customers fill out and respond to the survey. However, from the association of the type of product purchased by customers and the type of product in association rules, it can be said that recommendation system can provide relevant products to the customers.

Recommendation system can filter the candidates of email recipients from 926 customers to 14 customers. On the other word, filter on email marketing recipients in new system is 6514% more effective than previous system. The number of email marketing recipients affected the duration of email marketing delivery. In the previous system, Webeli.com needs 1650 seconds to deliver email marketing to 926 customers (all customers). The new system only needs 32 seconds to deliver email marketing to 14 customers (customers who have relevant interest with product combination in association rules). On the other words, duration of email marketing delivery in the new system is 5056% quicker than previous system. The number of email marketing recipient also affects significantly the email marketing cost. In the new system, email marketing costs 6517% cheaper than in the previous system.

Based on the explanation above, can be concluded that recommendation system can provide relevant products to the customers and it can optimize email marketing campaign by decreasing the number of email marketing recipients, duration of email delivery, and cost of email delivery.

For the future research, Webeli has to try to combine some of the recommendation systems to make the recommendations more accurate and integrate Webeli's system with Interspire Email Marketer to send email marketing automatically.

References

[1] Abidin, I. S. Z., Bakar, N. A. A., & Haseeb, M.. An empirical analysis of exports between Malaysia and TPP member countries: Evidence from a panel cointegration (FMOLS) model. Modern Applied Science, 8(6), 238,(2014).

[2] Abidin, I. S. Z., Bakar, N. A. A., & Haseeb, M.. Exploring Trade Relationship between Malaysia and the OIC Member Countries: A Panel Cointegration Approach Asian Journal of Scientific Research, 8(1), 107,(2015), (1995-2012).

[3] Abidin, I. S. Z., & Haseeb, M. Investigating Exports Performance between Malaysia and OIC Member Countries from Asian Social Science, 11(7), 11. (2015), 1997-2012.

[4] Abidin, I. S. Z., Haseeb, M., Azam, M., & Islam, R. Foreign direct investment, financial Development, international trade and energy consumption: Panel data evidence from selected ASEAN Countries. International Journal of Energy Economics and Policy, 5(3),(2015).

[5] Abidin, I. S. Z., Haseeb, M., & Islam, R. Regional Integration of the Association of Southeast Asian Nations Economic Community: An Analysis of Malaysia-Association of Southeast Asian Nations Exports. International Journal of Economics and Financial Issues, 6(2),(2016).

[6] Abidina, I. S. Z., Haseeb, M., & Jantan, M. D. Trans-Pacific Partnership (TPP) Agreement: Comparative Trade and Economic Analysis for Malaysia. The Social Sciences, 11(13), 3375-3380, (2016).

[7] Badrul, M. PENENTUAN STRATEGI MARKETING PENJUALAN PRODUK DENGAN ALGORITMA APRIORI. Konferensi Nasional Ilmu Sosial & Teknologi, 1(1),(2017).

[8] Georgieva, M. An introduction to email marketing. How to Execute & Measure Successful Email Marketing, HubSpot,[online] Available at: http://offers. hubspot. com/an-introduction-to-email-marketing,[Accessed 15 October 2016],(2012).

[9] Han, J., Pei, J., & Kamber, M. Data mining: concepts and techniques: Elsevier,(2011).

[10] Pramudiya, H. E., Handarkho, Y. D., & Rahayu, F. S. Pengimplementasian CRM Pada Pembangunan E-Commerce untuk Usaha Mikro Kecil Menengah (Studi Kasus: Dolanan Puzzle). Jurnal Buana Informatika, 6(4),(2015).

[11] Roiger, R. J. Data mining: a tutorial-based primer: CRC Press,(2017)

[12] Sharif, M. N. A., Ching, N. M., Bakri, A., & Zakaria, H. Using a priori algorithm for supporting an e-commerce system. Journal of Information Technology Impact, 5(3), 129-138,(2005).

[13] Suryanto, T. Dividend policy, information technology, accounting reporting to investor reaction and fraud prevention. International Journal of Economic Perspectives, 10(1), 138,(2016).

[14] Suryanto, T., & Ridwansyah, R. The Shariah financial accounting standards: How they prevent fraud in Islamic Banking. European Research Studies, 19(4), 140,(2016).

[15] Zulfikar, W. B., Wahana, A., Uriawan, W., & Lukman, N. Implementation of association rules with apriori algorithm for increasing the quality of promotion. Paper presented at the Cyber and IT Service Management, International Conference on,(2016).