Segmentation-Based Sequential Rules For Product Promotion Recommendations As Sales Strategy (Case Study: Dayra Store)

Dayan Ramly Ramadhan*[1], Nur Rokhman[2]
[1]Master Program of Computer Science; FMIPA UGM, Yogyakarta, Indonesia
[2]Department of Computer Science and Electronics; FMIPA UGM, Yogyakarta, Indonesia
e-mail: *[1]**d.ramly@mail.ugm.ac.id**, [2]nurrokhman@ugm.ac.id

Abstrak

Salah satu kendala pada promosi terletak pada tingginya biaya. Semakin penjual tidak mengetahui sasaran promosi, semakin besar biaya yang dikeluarkan. Dengan mengetahui segmen pelanggan yang pernah melakukan transaksi, serta perilaku transaksi barang yang dibeli secara berurutan, penjual dapat mempromosikan produk kepada konsumen potensial. Penelitian ini mengolah data transaksi penjualan Dayra store dari tahun 2017 hingga 2019 dan menggunakan metode segmented-based sequential rule yang menghasilkan pola sekuensial dari masing-masing segmen pelanggan untuk menjadi rekomendasi promosi produk. Pada metode tersebut diuji coba dengan membandingkan promosi produk bedasarkan rules dan promosi produk tanpa berdasarkan rules. Hasil promosi produk kemudian dilakukan pengujian hipotesis dan uji statistik. Berdasarkan hasil pengujian promosi produk, diperoleh rata-rata persentase transaksi promosi produk berdasarkan rules sebesar 2,622%, lebih tinggi lebih tinggi dibandingkan dengan promosi dengan produk terbaru dengan rata-rata persentase transaksi 0,315%. Dari hasil uji hipotesis pada masing-masing segmen diperoleh dari sampel dapat mendukung pernyataan bahwa promosi produk pada keseluruhan segmen berdasarkan rules dapat lebih efektif meningkatkan penjualan dibandingkan dengan promosi yang menggunakan produk terbaru tanpa menggunakan rekomendasi rules.

Kata kunci—*Data Mining, Sequential Rules, Rule-Growth, Segmentasi Pelanggan, RFM Model*

Abstract

One of the problems in the promotion is the high cost. Identifying the customer segments that have made transactions, sellers can promote better products to potential consumers. The segmentation of potential consumers can be integrated with the products that consumers tend to buy. The relationship can be found using pattern analysis using the Association Rule Mining (ARM) method. ARM will generate rule patterns from the old transaction data, and the rules can be used for recommendations. This study uses a segmented-based sequential rule method that generates sequential rules from each customer segment to become product promotion for potential consumers. The method was tested by comparing product promotions based on rules and product promotions without based on rules. Based on the test results, the average percentage of transaction from product promotion based on rules is 2,622%, higher than the promotion with the latest products with an average rate of transactions only 0,315%. The hypothesis in each segment obtained from the sample can support the statement that product promotion in all segments based on rules can be more effective in increasing sales compared to promotions that use the latest products without using rules recommendations.

Keywords—*Data Mining, Sequential Rules, Rule-Growth, Customer Segmentation, RFM Model*

1. INTRODUCTION

Marketing strategy for a business requires customer identification to maximize sales. Each customer tends to choose different products in different transactions. But the tendency to choose this product can be studied so that the patterns become new knowledge in determining sales strategies. Dayra store is one of the stores that sell their products online. The products sold by Dayra store are clothing materials used by Muslim women such as hijab, dress, and mukenah. In this type of product, consumers generally buy products at different times and transactions.

At present, the impediment to promotion lies in the high cost. The more the seller does not know the promotional goals, the higher the costs. An analysis is needed to make advertising to be more optimal [1]. By utilizing the forecast model, online businesses can minimize the cost of advertising aimed only at potential consumers. This model can be done by predicting the right product for a consumer from past consumer transactions [2].

Data Mining used is how to trace existing data to build a model, then make the model able to recognize other data patterns that are not in a stored database. In this case, customer segmentation is one of the first steps in creating a business model [3]. Customer segmentation divides customers into separate groups that have similar characteristics. The segmentation method has traditionally identified customers with several key variables such as demographics, psychographics, and behavioral variables, while the value-based segmentation method identifies customer groups based on the amount of revenue they produce and the age value of customers [4].

The results of the segmentation can be used in finding sales patterns. In finding patterns, one method in data mining is market basket analysis. In this method, finding a relationship between one item with another item occurs, so that it can know the criteria for a tendency to buy to make a real transaction. If at a sale with a physical store, managing buying tendencies are used by placing items that are often purchased by buyers simultaneously on adjacent shelves. However, online sales are used to carry out promotions to consumers.

In the Dayra store case study, the segmentation of potential consumers can be integrated into products that consumers tend to buy. Therefore, the relationship between consumer segmentation and product identification can be found using pattern analysis. Association Rule Mining (ARM) will produce a pattern of rules from the old transaction data, and the resulting rules can be used for recommendations [5]. By knowing the customer segments that have carried out transactions, as well as the transaction behavior of goods purchased sequentially, sellers can promote the product to potential consumers who have bought goods before. This study uses a segmented-based sequential rule method, which is a development method of the concept of sequential rule-based [6]. With the addition of customer segmentation, this research is expected to produce product recommendations based on sequential patterns that can provide better results.

2. METHODS

2. 1 Customer Segmentation

Segmentation is the process of dividing customers into clusters with customer loyalty categories to build marketing strategies. Customer segmentation divides the market into separate customer groups that have similar characteristics. The segmentation method has traditionally identified customers with several key variables such as demographics, psychographics, and behavioral variables, while the value-based segmentation method identifies customer groups based on the amount of revenue they produce and the age value of customers [4]. Many metrics can be used to measure customer value, one of which is Customer Lifetime Value (CLV).

Customer lifetime value (CLV) usually used to identify profitable customers and develop strategies to target customers [7].

Several ways can be used to identify the value of customer life, including Recency, Frequency, and Monetary (RFM) analysis. RFM model analyzes customer behavior and then predicts customer loyalty based on the behavior of customers who transact recorded in the customer transaction database [8]. Customers are segmented to various target markets according to their RFM value. The advantage of the RFM model lies in its relevance while operating on several observable and objective variables. The RFM value is defined as follows (1) R (Recency): the time interval between the last purchase and the current time. The lower value corresponds to the higher probability that the customer will make a repeat purchase. (2) F (Frequency): the number of purchases made within a certain period, higher frequencies indicate stronger loyalty. (3) M (Monetary): the amount of money spent during a specific period.

The higher the value of R and F, the more customers tend to make repeat transactions. While the higher the value of M indicates that the tendency of customers to buy the company's products or services. RFM is an effective attribute for segmenting customers [9]. Customer segmentation divided into six characteristics based on the RFM values shown in Table 1.

Table 1 Customer characteristics

Customer Segment	Characteristics
Superstar	a. Customers with high loyalty. b. Customers with the highest monetary value. c. Customers with the highest frequency. d. Customers with the highest recency.
Golden Customer	a. Customers have high recency value. b. Customers have high-frequency value. c. Customers have the average monetary value of the entire customer.
Typical Customer	a. Customers have the average monetary value of the entire customer. b. Customers have a moderate frequency.
Occational Customer	a. The second-lowest frequency value after the dormant customer. b. The lowest recency value
Everyday Shopper	a. Customers have higher frequency than dormant customers. b. Customers have moderate to low monetary value.
Dormant Customer	a. Customers have low frequency and monetary. b. The lowest recency value.

2. 2 Sequential Rule Mining

Sequential rule mining is one method that applies data mining techniques to sequential databases to find correlations that exist between lists of events/items and find frequent frequencies as patterns [10]. The main difference with association rule mining is the concept of time or sequence in sequences. A sequential rule $X \rightarrow Y$ can mean X followed by Y, while an association rule $X \rightarrow Y$ can mean that if X appears, then Y will also appear and can occur at one time [11].

The application of sequence data is very commonly used, which can be found in many domains such as bioinformatics (DNA sequence), the sequence of visitor clicks on the website, learner behavior in e-learning, order of customer purchase lists, and sentences in a text. All transactions from the same customer can be grouped together, the transaction then sorted by the time ascending (increasing) into a customer sequence.

Sequential rule mining is the search for all repeated sequences, i.e., sequences whose frequency of occurrence is greater than the minimum-support [12]. A sequence consists of a series of elements, and each element consists of several ordered items. A sequential rule $X \rightarrow Y$

is a sequential relationship between two sets of items X and Y, with X not sequential and Y, not sequential. Support for the rule X → Y is the number of sequences containing all X items before all items from Y divided by the number of sequences in the database. The confidence of the rule is the number of sequences containing all items X before all Y items, divided by the number of sequences containing items in X.

Table 2 *Sequential rules*

ID	Rule	Support	Confidence
r1	{*jilbab navilla*} → {*jilbab anak nagita*}	2	0.67
r2	{*jilbab navilla*} → {*jilbab narina*}, {*jilbab anak nagita*}	1	0.33
	{*jilbab husna*} → {*gamis cassandra*}		
r3	{*jilbab anak nagita*} → {*jilbab aira*}	1	0.33
r4	{*jilbab narina, jilbab navilla, jilbab anak nagita*} →	1	1
r5	{*jilbab aira*}	1	1
…	…	…	…

2. 3 Segmentation-based Sequential Rule (SSR)

The segmented-based sequential rule (SSR) method is a development of the sequential rule-based (SR) method by improving the quality of recommendations based on customer groups [13]. The SSR method extracts sequential rules from each customer group and provides recommendations based on the target customer group. Customer-based recommendations offer more knowledge about customer behavior in the future, such as how often the customer will buy, and how much the purchase will be.

The main advantage of customer-based recommendations is adopting different marketing strategies for different customer segments. In addition, grouping customers into different groups will improve the quality of recommendations and help decision-makers identify market segments more clearly and develop more effective strategies [8].

The SSR method considers the order of customer purchases in which the SR method is ignored. This method consists of two phases: the model formation phase and the recommendation phase [13]. The model formation phase includes the process of customer segmentation and sequential-rule search for each customer segment in the segmentation process using RFM (Recency, Frequency, and Monetary) analysis. Then the results of the segmentation continued on the recommendation process using sequential rule mining. The sequential rule mining process implements the Rule-Growth algorithm to focus on customer purchase patterns in each segment. By knowing all the possible patterns of each customer segment's rules, we can provide appropriate product recommendations as product promotion materials. The following is a general description of the segmented-based sequential rule (SSR) method in research.
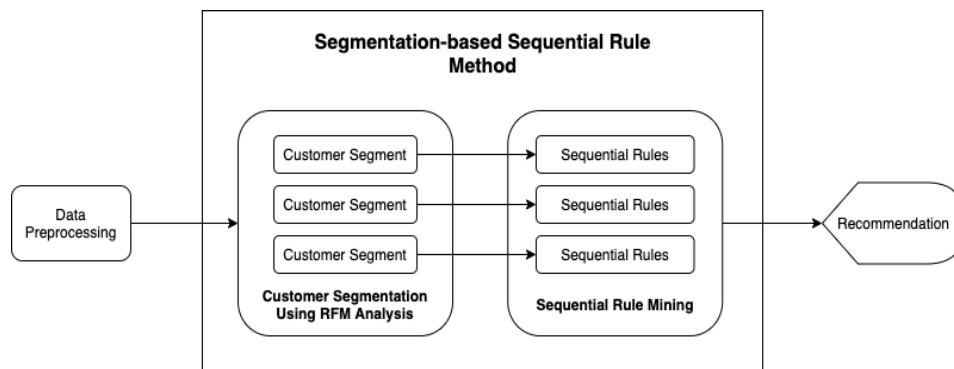

Figure 1 *Segmented-based sequential rule* method

2. 3.1 Customer Segmentation Using RFM

RFM value is derived from transaction data based on customer purchases that have been extracted and normalized. If the customer shows the same buying behavior or made a similar purchase in the previous period, the customer likely has the same RFM value in this period. Customers with the same RFM values are grouped into customer segments to model and make recommendations. The variables needed for customer segmentation are the order number, the number of products ordered the time of payment, and the total payment. Recency, frequency, and monetary values are divided into four parts (quartiles), which are equal to values 1, 2, 3, and 4.

The recency value is calculated based on the interval of the last transaction date of all transactions with the date of the last transaction of each customer. Value 4 is given to customers with the most recent transaction date and value 1 for customers with the farthest transaction date in the past. The frequency value is calculated from the number of transactions made by the customer. Customers who frequently trade are given a value of 4, while customers who rarely trade are given a value of 1. Monetary value is calculated from the total amount of money that has been spent by customers. Customers who spend a lot of money have a high monetary value, which is a value of 4. At the same time, customers who spend less money have a low monetary value, which is a value of 1.

Each customer's RFM value is then added together to get the overall RFM value. Customers with the same total RFM value can be combined into one group that is defined as one segment customer based on the following Table 3.

Table 3 Customer segmentation criteria from RFM value

Customer Segment	RFM Value
Superstar	12
Golden customer	10 - 11
Typical customer	8 – 9
Occational customer	6 - 7
Everyday shopper	4 - 5
Dormant customer	<4

2. 3.1 Rule-Growth Algorithm

The rule-Growth algorithm is an alternative method in mining sequential rule patterns to predict rules with their confidence value. Rules are defined as conditions that, when several items occur, other items will appear after that in a certain sequential data set. This mechanism also helps us predict patterns [10]. In addition, study the Rule-Growth algorithm's performance and deduce efficiency in finding sequential rule patterns. The Rule-Growth algorithm is very efficient, faster than CMDeo and CMRules [14].

The Rule-Growth has the main task of finding valid rules on several consecutive sets of items. Two variables must be determined, the value of confidence, and the value of support [10]. Following are the steps carried out in the Rule-Growth algorithm:

- First, scan the database to find items that frequently appear, for example. {s1, s2, s3, s4, ...}.
- For each pair of items, try to make a rule with only these two items, for example, {s1} => {s2}.
- Then, find a larger rule by re-scanning the database repeatedly and adding one item at a time to the left (antecedence) or right (consequent) of each rule (left and right expansions). For example {s1, s3} => {s2}, then {s1, s3, s4} => {s2}.
- To determine items that can be expanded on the left in the rule A => B: scan the sequence containing the rule and the items that appear at least meet the minimum support (minsup) sequence that occurred before the last appearance B.

- To determine items that can be expanded on the right in the rule A => B: scan the sequence containing the rule and the items that appear at least meet the minimum support (minsup) sequence that occurred before the first appearance B.
- To avoid generating the same rule twice: Add only items to the left/right of the rules if the item is larger than all items that are already on the left/right. Do not allow expansion to the left after expanding on the right, but it is possible to expand on the right side after expansion to the left.
- Test the validity of the rules.

2. 4 Lift Ratio

One way to see whether the rules are strong is to compare it with benchmark values. Benchmark confidence compares the number of transactions with items in consequence with the number of transactions in the dataset. The value of benchmark confidence can be calculated using the formula shown by the equation below [15].

$$Benchmark\ Confidence\ (A \rightarrow B) = \frac{Nc}{N} \tag{1}$$

This value is compared with the value of confidence, called the lift ratio. So the lift ratio is a comparison between confidence for a rule divided by benchmark confidence, which is assumed to be consequent and antecedent independent of each other. The equation below shows the calculation for obtaining a lift ratio [15].

$$Lift\ Ratio\ (A \rightarrow B) = \frac{Confidence\ (A \rightarrow B)}{Benchmark\ Confidence\ (A \rightarrow B)} \tag{2}$$

If the lift ratio value is greater than 1, this shows the rule is strong. The higher the lift ratio, the greater the strength of the rules. The calculation of lift ratio values in sequential rule mining can be adapted from the calculation of lift ratio values in association rule mining with a slight modification in the algorithm.

2. 5 Hypothesis Test

The hypothesis is a temporary answer to the formulation of the research problem [16]. The hypothesis needs to be tested to find out whether the product recommendations based on the rules obtained have an impact on sales. The rules impact if product promotion based on rules further increases sales, which can be seen from the transactions that occur. The promotion will be compared with the latest products that have the same value, and sales have occurred on previous transactions.

Hypothesis testing the proportion of two populations is testing two proportions, each of which is derived from two different and independent populations. Testing two proportions is used when comparing whether the proportion in the first population is smaller, equal to or greater than the proportion in the second population. The following equation shows the test statistics used in the proportion test for two populations.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\ (1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}} \tag{3}$$

Where

$$\hat{p}_1 = \frac{x_1}{n_1} \tag{4}$$

$$\hat{p}_2 = \frac{x_2}{n_2} \qquad\qquad (5)$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \qquad\qquad (6)$$

Notes :
$\hat{p}_1$ = The proportion in sample 1
$\hat{p}_2$ = The proportion in sample 2
$\bar{p}$ = Combined proportion
x_1 = The number of successful transactions in sample 1
x_2 = The number of successful transactions in sample 2
n_1 = Number of samples 1
n_2 = Number of samples 2

3. RESULTS AND DISCUSSION

The concept of Segmentation-Based Sequential Rules is carried out on sales data that have been collected and implemented on a system to produce rules that represent each segment. The system implementation uses the python programming language. The process that occurs in the system consists of pre-processing data, customer segmentation, Segmentation-Based Sequential Rules mining, and testing. The testing process uses a comparison of minimum support to produce rules for each segment. The rules obtained are then selected with the highest lift ratio and the stock that is still available for promotion to customers.

The effectiveness of segmented-based sequential rule (SSR) methods tested on six customer segments for generating rules. The testing process will be applied in the different minimum support values so that rules are found in each customer segment. Tests are carried out to compare pattern searches with minimum support of 4%, 3%, and 2%. The results of testing the rules can be seen in table 4.

Table 4 Rules testing results

Customer Segment	Rules		
	MinSup 4%	*MinSup* 3%	*MinSup* 2%
Superstar	0	12	853
Golden Customer	3	16906	15442
Typical Customer	0	0	762
Occational Customer	0	0	277
Everyday Shopper	0	774	767
Dormant Customer	37	48	37

From the test results based on the minimum value of support, the smaller the minimum support value, the more rules are produced. This result happens because frequent itemset limitation decreases so that not all itemset can be used as rules and combined in the next sequence. The following testing is done by giving promotions to each segment using a minimum support limit of 2%. The choice of rules tested at this minimum support is because all segments get rules at this limit.

The rules obtained are used to design marketing campaigns by using discount coupons for certain item combinations. Promotion is carried out for nine days in the early weeks of April 2020. According to each customer segment, the tools used in the promotion are facebook business and advertised with target customers. The selected products that are promoted are based on the results of the rules in each segment, with the largest lift ratio conditions and available stock. The results of product promotions then compared with product promotions

using the latest products. The selected product is the newest product chosen randomly, and it should have sales transactions before. The results of product promotion can be seen in Table 5 .

Table 5 Product promotion results

Segments	Using rules			Latests product without rules		
	Potential Buyers	Transactions	Percentage Transaction	Potential Buyers	Transactions	Percentage Transaction
Superstar	147	7	4,762 %	95	1	1,053 %
Golden Customer	97	2	2,062 %	86	0	0
Typical Customer	131	3	2,290 %	119	1	0,840 %
Occational Customer	134	1	0,746 %	121	0	0
Everyday Shopper	67	2	2,985 %	86	0	0
Dormant Customer	104	3	2,885 %	91	0	0
	Average		2,622 %	Average		0,315 %

Based on the product promotion results in table 5, the column for potential buyers is the number of potential buyers interested in promotion. The transaction is the number of interested buyers who make a purchase. From the data above, a hypothesis test using Z score with the value for $\alpha = 0.005$ is -1.645. The testing process used a proportion test of two populations, then the statistic test used z test which applied on each customer segment. Z scores value of all segments can be seen in the following table 6.

Table 6 Z scores table each segments

Segments	Z score
Superstar	1,5760
Golden Customer	1,3390
Typical Customer	0,9124
Occational Customer	0,9521
Everyday Shopper	1,6128

All segments have Z scores located in the acceptance area of Ho. This result means that the information obtained from the sample can support the statement that product promotion in each segment based on rules can increase sales compared to promotions with the latest products that do not use rules recommendations.

4. CONCLUSIONS

Customer segmentation with RFM values based on transaction data produces six customer segments. There are 88 customers in the superstar segment, 33 customers in the Golden Customer segment, 44 customers in the Typical Customer segment, 37 customers in the Occasional Customer segment, 26 customers in the Everyday Shopper segment, 19 customers in the Dormant Customer segment. From each customer segment, rules found at a minimum support of 2%. The results are 853 rules in the superstar segment, 15442 rules in the Golden Customer segment, 762 rules in the Typical Customer segment, 277 rules in the Occasional Customer segment, 767 rules in the Everyday Shopper segment, 37 rules in the Dormant Customer segment.

The results of product promotions in each segment obtained the percentage of transactions that occur in the promotion based on rules is higher than the promotion using the latest products without using the rules recommendations. The result is an average percentage of transactions of 2.622% in promotions based on rules and 0.315% in promotions that are not

based on rules. The hypothesis testing results in each segment obtained from the sample can support the statement that the promotion of products in all segments based on the rules can be more effective in increasing sales compared to promotion using the latest products without using rules recommendations.

5. SUGGESTION

Develop a Segmentation-based Sequential Rule with a combination of behavioral and demographic data for customers to get better product recommendations. Development can also be done with other classification algorithms to get more detailed classification results.

REFERENCES

[1]   J. Zhang and J. Li, "Retail Commodity Sale Forecast Model Based on Data Mining," *Int. Conf. Intell. Netw. Collab. Syst.*, pp. 307–310, 2016.

[2]   Y. Liu and Y. Guan, "Application in Market Basket Analysis Based on FP-growth Algorithm," *WRI World Congr. Comput. Sci. Inf. Eng.*, vol. 4, pp. 112–115, 2009.

[3]   B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *Jutei*, vol. 2, no. 2, pp. 23–32, 2018, doi: 10.21460/jutei.2017.21.76.

[4]   D. R. Liu and Y. Y. Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Inf. Manag.*, vol. 42, no. 3, pp. 387–400, 2005, doi: 10.1016/j.im.2004.01.008.

[5]   D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems An Introduction*. New York: Cambridge University Press, 2011.

[6]   D. R. Liu, C. H. Lai, and W. J. Lee, "A hybrid of sequential rules and collaborative filtering for product recommendation," *Inf. Sci. (Ny).*, vol. 179, no. 20, pp. 3505–3519, 2009, doi: 10.1016/j.ins.2009.06.004.

[7]   M. S. Kahreh, M. Tive, A. Babania, and M. Hesan, "Analyzing the Applications of Customer Lifetime Value (CLV) based on Benefit Segmentation for the Banking Sector," *Procedia - Soc. Behav. Sci.*, vol. 109, no. Clv, pp. 590–594, 2014, doi: 10.1016/j.sbspro.2013.12.511.   [Online].   Available: http://linkinghub.elsevier.com/retrieve/pii/S1877042813051434

[8]   D. Birant, "Data Mining Using RFM Analysis," *Knowledge-Oriented Appl. Data Min.*, no. iii, 2011, doi: 10.5772/13683.

[9]   J. Wu and Z. Lin, "Research on customer segmentation model by clustering," *ACM Int. Conf. Proceeding Ser.*, vol. 113, pp. 316–318, 2005, doi: 10.1145/1089551.1089610.

[10]   Y. W. Trio Pramono and Suhardi, "Anomaly-based intrusion detection and prevention system on website usage using rule-growth sequential pattern analysis: Case study: Statistics of Indonesia (BPS) website," in *Proceedings - 2014 International Conference on Advanced Informatics: Concept, Theory and Application, ICAICTA 2014*, 2015, doi: 10.1109/ICAICTA.2014.7005941.

[11]   P. Fournier-Viger, C. W. Wu, V. S. Tseng, L. Cao, and R. Nkambou, "Mining Partially-Ordered Sequential Rules Common to Multiple Sequences," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2203–2216, 2015, doi: 10.1109/TKDE.2015.2405509.

[12]   R. Juliastio and D. Gunawan, "Sequential Pattern Mining Dengan Spade Untuk Prediksi Pembelian Spare Part Dan Aksesoris Komputer Pada Kedatangan Kembali Konsumen," pp. 314–325, 2015.

[13]   D. R. Liu, C. H. Lai, and W. J. Lee, "A hybrid of sequential rules and collaborative filtering for product recommendation," *Inf. Sci. (Ny).*, vol. 179, no. 20, pp. 3505–3519, 2009,   doi: 10.1016/j.ins.2009.06.004.   [Online].   Available:

http://dx.doi.org/10.1016/j.ins.2009.06.004

[14]   P. Fournier-viger, R. Nkambou, and V. S. Tseng, "RuleGrowth : Mining Sequential Rules Common to Several Sequences by Pattern-Growth," *Symp. Appl. Comput.*, 2011, doi: 10.1145/1982185.1982394.

[15]   B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.

[16]   Sugiyono, *Statistika Untuk Penelitian*. Bandung: CV Alfabeta, 2007.