ISSN 2614-0047

Sequential pattern mining to support customer relationship management at beauty clinics

Esther Irawati Setiawan [a,1,*], Valerynta Natalie [a,2], Joan Santoso [a,3], Kimiya Fujisawa [b,4]

[a] Institut Sains dan Teknologi Terpadu Surabaya, Indonesia
[b] Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology, Tokyo
[1] esther@stts.edu; [2] valerynta.natalie@gmail.com; [3] joan@stts.edu; [4] fujisawa@stf.teu.ac.jp
* corresponding author

ARTICLE INFO

Article history
Received October 30, 2022
Revised November 27, 2022
Accepted December 3, 2022

Keywords
Data Mining
Generalized sequential pattern
Sequential pattern mining

ABSTRACT

The increasing competition for beauty clinics makes management need to think of methods to survive in this competition. For that, the company needs to improve CRM in its customer service. Customer Relationship Management is a series of activities managed to understand better, attract attention, and maintain loyalty. Sequential Pattern Mining is one of the data mining techniques that is useful for finding sequential patterns / sequences of a set of items. The algorithm that is used is the Generalized Sequential Pattern (GSP). GSP performs candidate generation and supports counting processes, that are, the union of L1-k with itself, which generates a candidate sequence that cannot exist as a twin candidate after that deletion candidate who does not meet the minimum support. While carrying out the process through existing data, it is also carried out increasing the number of supports from the included candidates in data sequences. The output to be produced by the program are all frequent itemsets that satisfy minimum support in the form of rules. Sales transaction data will be processed by using the Generalized Sequential Pattern algorithm so that it can produce a rule, namely the purchase order that meets the minimum support. The result of the rule used by management to support enterprise CRM activities such as acquiring new customers, increasing the profits from existing customers, and retaining existing customers.

1. Introduction

The customer is the leading actor in the sales transaction that will generate profit for the company. Competition in the business world is getting intense. Every company is competing to attract attention from the market, and customers increasingly have more choices in fulfillment of their needs. Every company's challenge is making customers visit and stay loyal [1], [2]. Often companies only focus on products sold and ignore services that cannot be ignored, this is very closely related to Customer Relationship Management (CRM), where a series of activities is managed to understand better, attract attention, and maintain the loyalty of customers [3], [4], this is the basis for doing this research.

By utilizing Sequential Pattern Mining, this research is expected to help company management parties take steps to meet customer satisfaction. Sequential Pattern Mining is one of the valuable data mining techniques to find the sequential pattern of a group of items [5]. This study uses the Generalized Sequential Pattern (GSP) algorithm to find rules of transactions that will later support CRM in the company. The Association Mining method could be combined with the Generalized Sequential Pattern (GSP) Algorithm to identify sequences or patterns of attributes that frequently occur together, to generate recommendations for movies to watch after a previous film has ended [6]. The researchers found that the GSP algorithm effectively identified association rules and sequential

pattern rules based on movie transaction data, which could be used to recommend films and increase audience interest in watching movies.

The Generalized Sequential Pattern (GSP) algorithm can identify user behavior patterns in each transaction, revealing relationships or associations between books that are requested simultaneously or sequentially [7]. Through the algorithm's calculations, a total of 295 frequent sequences consisting of three sequence patterns have been identified based on a minimum support threshold of 0.53% or a minimum number of two borrowed books. In a recent study [8], researchers used a combination of qualitative textual analysis, human-based content analysis, and machine learning techniques to examine user-generated content (UGC) on social media, focusing on Dove's "Campaign for Real Beauty" as a case study. The study outlines a six-step analysis procedure that includes identifying topics through qualitative analysis, generating labeled data through human coding, preprocessing data, evaluating machine learning classifiers, classifying unlabeled data, and conducting research. The findings of this study have significant methodological implications for advertising scholars and practitioners, particularly in the beauty industry, and can be applied to similar research studies.

CRM could utilize Association rule mining and sequential pattern mining techniques to provide recommendations to customer service [9]. CRM is built based on mobile and is able to provide effective services and recommendations for customers. From this study, it was found that CRM is able to maintain the quality of the company's relationship with customers through the utilization of information about the customer. A recent study has investigated the patterns of structural changes in customer segments and proposed a new approach that combines clustering and sequential rule mining techniques. To test the proposed method, the researchers applied it to customer data from a telecommunication service provider, demonstrating its effectiveness in this field. One interesting finding was the identification of a group of customers who exhibited dynamic behavior that caused structural changes, and the researchers labeled this group as "structure breakers." The insights gained from this study can be helpful for marketing managers at the telecommunication company, as they can leverage these results to refine their marketing strategies and improve their decision-making processes. This new approach could also be applied to other organizations to analyze patterns of structural changes in customer segments [10].

2. Method

2.1. Related Works

2.1.1. Sequential Pattern Mining (SPM) [11], [12]

One of the data mining techniques can find patterns in order of a set of items that will result in output in the form of rules. According to [13], the Generalized Sequential Pattern Mining (GSP) algorithm is an algorithm that can process and find all existing sequential and non-sequential patterns. Input from SPM is a data sequence, a collection of data sequences. Each data sequence is a list of transactions consisting of items.

In general, each transaction is associated with transaction time. No data sequence has more than one transaction with the same transaction time and uses transaction time as a transaction identifier, where in this case, the quantity of an item in a transaction is not taken into account. Agrawal and Srikant first introduced Sequential Pattern Mining. According to [14], GSP algorithms are generally viewed as the first traversal area algorithm that finds all sequence that frequently occurs by passing several data. The SPM algorithm is divided into two main methods, which are:

- A priori-based, consisting of the GSP algorithm, which is a Sequential Pattern Mining method with a horizontal format, and the Sequential Pattern Discovery using Equivalent Class (SPADE) algorithm by adopting a vertical format SPM.
- Projection-based, consisting of the Freespan algorithm and Prefixspan, which applies a division pattern and a series of strategies for the efficiency of Sequential Pattern Mining.

2.1.2. Generalized Sequential Patterns (GSP)

The GSP [15] algorithm works by analyzing existing data to identify sequential patterns. It involves multiple phases, each determining the support of items in the data. The support is the number

of data sequences that contain the items. The algorithm then identifies items that meet the minimum support level and are therefore considered frequent.

Each frequent item produces a frequent sequence, with the first consisting only of the item. In each phase, the algorithm starts with a set of potential candidates, frequent sequences from the previous phase. These candidates produce new potential frequent sequences, which must have more than one item in ordinary with the original candidate sequence. The algorithm determines the support of each candidate sequence as it progresses through the data. At the end of each phase, the algorithm identifies which candidate sequences are frequent and adds them to the list of candidates for the next phase. The process continues until no more frequent sequences can be generated or no more candidate sequences are left to analyze.

Candidate Generation is the stage where the set of all frequent(k-1)-itemset F_(k-1) found on the pass to-(k-1) is used to generate candidate itemset Ck. The Join Phase generates the candidate sequence by doing a join process or merging of L1−k with itself. Prunes Phase deletes candidate sequences that do not meet the specified minimum support. On Counting Candidates, while doing the process through existing data, additional amounts are also made to support the candidate included in the data sequences. The GSP algorithm can be seen in Fig.1 [16].

Generalized Sequential Pattern

L_1= {large 1-sequences};
for (k=2; $L_k \neq \emptyset$; k++) do
begin
C_k= new candidates generated from L_{k-1} **foreach** customer-sequences in the database **do**
increment the count of all candidates in C that are contained in C.
L_k= candidates in C_k with minimum support.
end

Generate L_{k-1} Candidate

insert into C_k
select $p.litemset_1, \ldots, p.litemset_{k-1}, q.litemset_{k-1}$
from $L_{k-1\,p}$, $L_{k-1\,q}$
where $p.litemset_1 = q.litemset_1, \ldots, p.litemset_{k-2} = q.litemset_{k-2}$;

Rule Generation

RuleGen(F, min_conf);
for all frequent sequences $\beta \in F$ **do**
for all subsequences $\alpha \leftarrow \beta$ **do**
conf = fr(β)/fr(α);
if (conf ≥ min_conf) **then**
output the rule $\alpha \Rightarrow \beta$, and conf

Fig. 1.GSP algorithm

2.1.3. Customer Relationship Management (CRM) [17]

CRM [18] is a type of management that addresses theory on handling relationships between companies with its customers with the aim of increasing the value company in the eyes of the customer. Many CRM studies say that not all customers have an equal contribution to the business; therefore, to maximize business profits, it is necessary to evaluate each customer's value before designing a marketing strategy.

The main goal of every company for CRM is to analyze customer value and to improve customer retention rate. Customer retention is a process to retain old customers. To achieve the goal in this way, the company can incorporate the right constraints in Sequential Pattern Mining over time on existing transactions. Sequential buying patterns from customers help in determining customers' next buying behavior. Therefore, if the constraints are selected correctly, then customer value analysis and customer retention, the two important pillars of CRM, can be achieved.

Sequential Pattern Mining is one of the data mining techniques that is useful for finding sequential patterns of a set of items. This research using the Generalized Sequential Pattern Algorithm (GSP) to find future transaction rules will support CRM in the company [19], [20]. 17 most important parameters from the perspective of customer value as show in Fig. 2.

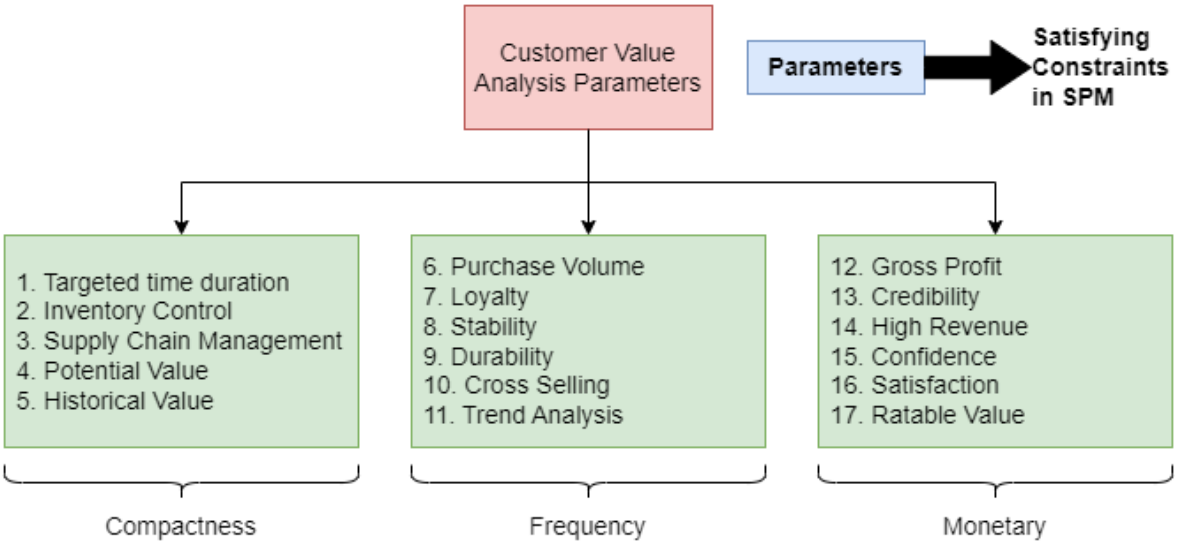

Fig. 2. Customer Value Analysis Parameter

There are 17 most important parameters from the perspective of customer, It will be discussed how constraints selection can be made to meet certain objectives of customer value. The 17 parameters are divided into 3 groups, namely Compactness, Frequency, and Monetary.

- Compactness

  Compactness constraint is important to use on sequential patterns due to customer buying habits varies from time to time. So, applying this technique can not only get new customers but also increase the customer's subsequent purchases which exist as show in Fig. 3.

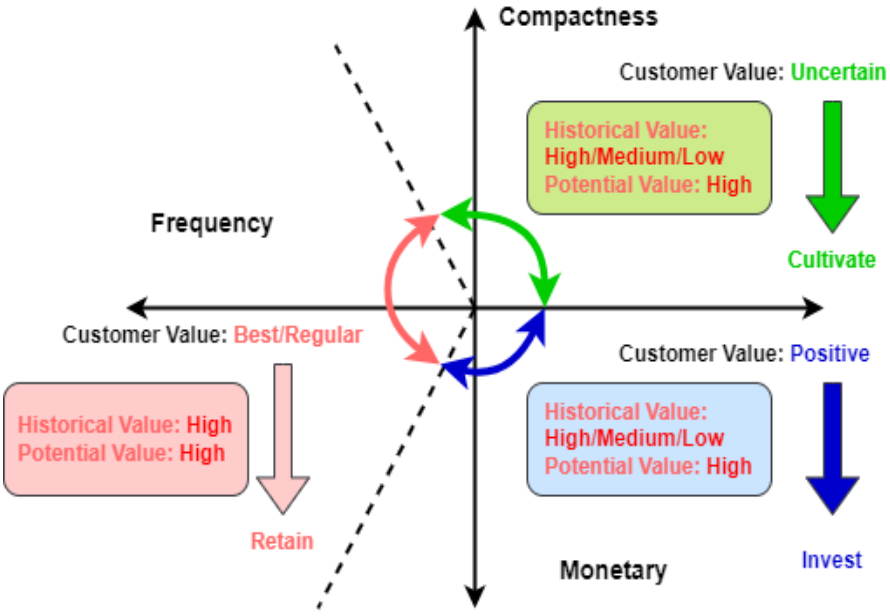

Fig. 3. Customer Value by Constraint Based on Sequential Pattern

- Frequency

  To focus on existing customers and increase retention rates, as show in Fig. 4.

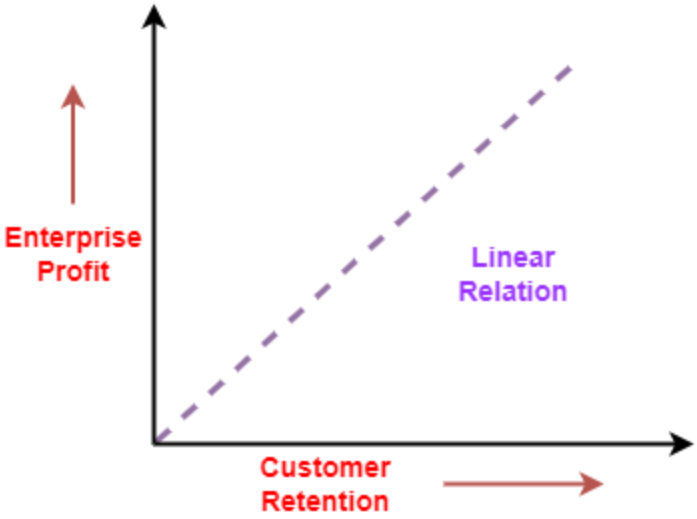

Fig. 4. Relationship between Customer Retention and Enterprise Profit

- Monetary

Monetary shows the amount of customer money issued for the product. Company fee for getting new customers is increasing. Lots of company agrees that the cost of obtaining new customers is 6-8 times the cost for retain existing customers. Therefore, it is clear that the company should give more attention to retaining existing customers.

2.2. System Architecture

After the data is ready, then the mining process is carried out in accordance with the stages in the GSP algorithm. It is the mining process that will produce the purchase rule/pattern from customers. With the rules that have been generated by the program, then the management can take concrete steps to carry out the CRM process on the customer. The Fig. 5 below is the overall block diagram of the system.

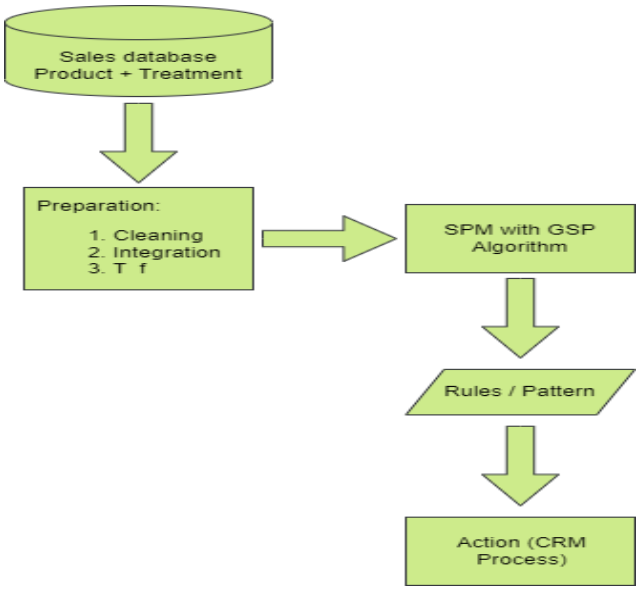

Fig. 5. Block Diagram

The last phase of Fig. 5 is Action (CRMProcess), namely the steps taken by the clinic based on the rules that the program has generated.

2.3. Application of CRM

The program was created using Microsoft Visual Studio with the C# programming language as show in Fig. 6.

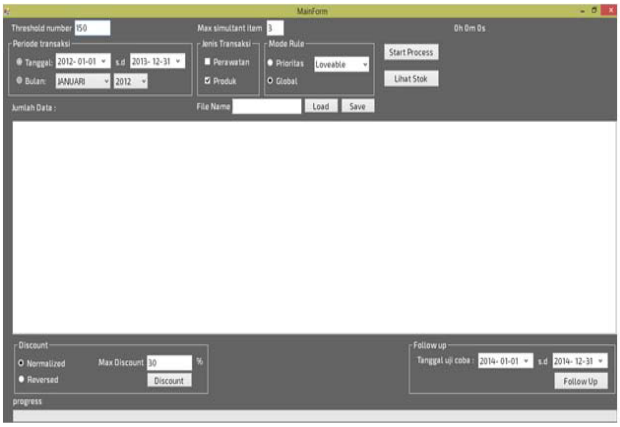

Fig. 6. Program Page Interface

In the program, there are several inputs for processing transaction data, namely:

- Transaction Periods can be selected by month or date. The type of transaction can also be selected according to the need for the rules to be generated. There are two types of transactions, namely Product and Care sales.
- There are also inputs for Minimum support which can be input as a percentage of the total number of transactions and the total desired transaction.
- Rule Mode consists of Priority and Global. Priority contains customer-level categories, while Global is all customers.
- View Stock is a button to display the last Stock of the product.
- Discount consists of numbers and options discount calculation input, namely Normalized and reversed. For example, a Discount normalized 30% is to give a maximum discount of 30% starting from the type of product/treatment with the most sales, whereas if selected, Reversed will start with the fewest sales.
- Follow Up will display a list of customers who have the potential to make transactions according to the results rule based on the order of previous purchase periods.

After the iteration process, the output generated is a series of rules as show in Fig. 7.

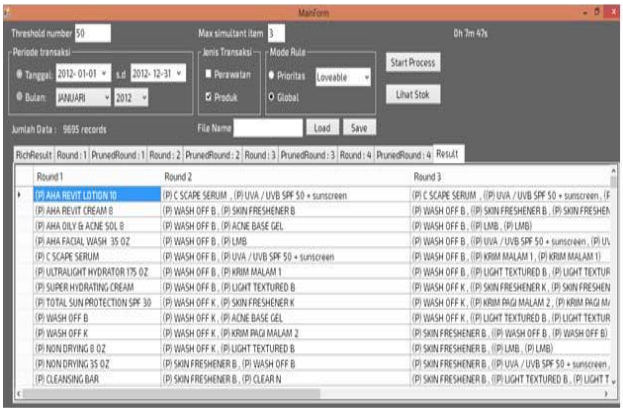

Fig. 7. Last Iteration Rules

Implementation will be linked to the discussion carried out in Chapter 2, namely in terms of the CRM phase, and on CRM. In general, the CRM phase is divided into three parts, namely:

- Acquire new customers (Acquire)

  One way to invite new customers is by promotion. Promotion can be customized with many factors, such as seasonal factors, trends, and specific events—for example, a promo to celebrate Valentine's Day. Then can perform a search in the February period of last year, and a pattern of a

purchase order will be found that can be used to design attractive promos and be liked by many people.

- Increase profits from existing customers (Enhance)

  Ways that can be done to improve benefits from existing customers can be through up-selling and cross-selling.

- Retain profitable customers (Retain)

  This research focuses on Retain by offering what specific customers need, not what market customers need. Application on the program is by taking advantage of the Follow Up feature, namely contacting or offering products/treatment recommendations to customers by the results of the rules and what has been purchased by the customer.

The things above can be applied to companies because customers will feel treated personally (because giving promos is not the same for everyone, depending on the customer's transaction history). Personal treatment will increase customer loyalty to the company [21].

2.4. Testing

- Trial Results Look Up

  Testing is done by matching the results of the rule from the training data process in the period between 2012-2013 with transactions that occurred in 2014 (testing data) as in Fig. 8.

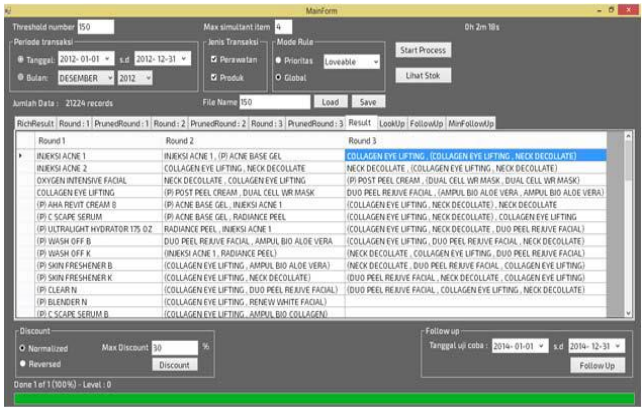

Fig. 8. Rule Testing

When the rule has been formed, then the program displayed as Fig. 8 can be clicked on one of the existing rules. Then the program will calculate on the testing data (2014 period) how many sequences are in accordance with the selected rule.

- Rule Testing

  The test was carried out with several different parameters. The test results are presented in tabular form in Table 1 below.

Table.1 Rule Testing Results

Year	Threshold	Max Simultan	Type (Transaction Total)	Time	Iteration
2012	100	3	Product, Care (21224)	15m 3s	3
2012	150	3	Product, Care (21224)	13m 20s	3
2012	200	3	Product, Care (21224)	12m 19s	2
2012	100	4	Product, Care (21224)	22m 5s	3
2012	150	4	Product, Care (21224)	22m 50s	3

3. Conclusion

From the results of this study, conclusions can be drawn, that the result of the resulting rule can be proven through 2014 trial data, that if there is a sequence of purchases in 2012 (training data) then the purchase order is also contained in the 2014 transaction (test data). From the analysis of the test results data then the test results with different thresholds can be concluded that with the same number of transactions, the time required and the rules generated are inversely proportional to the threshold value which is given. The greater the threshold value is given, the shorter the processing time and the resulting rules are also less, and vice versa.

The result of the rule can be used by management to support companies' CRM activities like getting new customers (Acquire), increasing the profits of existing customers (Enhance), and maintaining existing customers (Retain). The time required to process transaction data and generate rules is determined by the number of transactions, and the number of transactions depends on the length of the selected period. The longer the selected period and the smaller the threshold value becomes, the longer the processing time will be.

References

[1] R. Kalakota, M. Robinson, and D. Tapscott, *E-business 2.0: Roadmap for Success*, vol. 11, pp. 544. Addison-Wesley Boston, 2001. [Online]. Available: https://marmamun.gov.np/.

[2] S. Wilde, *Customer Knowledge Management: improving customer relationship through knowledge application*. Springer Science \& Business Media, 2011, doi: 10.1007/978-3-642-16475-0.

[3] R. J. Baran, R. J. Galka, and D. P. Strunk, *Principles of customer relationship management*. Cengage Learning, 2008.

[4] L. E. Herman, S. Sulhaini, and N. Farida, "Electronic Customer Relationship Management and Company Performance: Exploring the Product Innovativeness Development," *J. Relatsh. Mark.*, vol. 20, no. 1, pp. 1–19, Jan. 2021, doi: 10.1080/15332667.2019.1688600.

[5] C.-C. Chen, H.-H. Shuai, and M.-S. Chen, "Distributed and scalable sequential pattern mining through stream processing," *Knowl. Inf. Syst.*, vol. 53, no. 2, pp. 365–390, Nov. 2017, doi: 10.1007/s10115-017-1037-1.

[6] A. R. Andriyan, D. M. Rochma, M. N. Mudyawati, M. Jannah, S. L. D. Agustini, and A. A. Nugraha, "Determination of Film Recommendations using the Generalized Sequence Pattern (GSP) Association Method," in *Gunung Djati Conference Series*, 2021, vol. 3, pp. 7–11, [Online]. Available: https://conferences.uinsgd.ac.id/index.php/gdcs/article/view/89.

[7] T. Astuti and L. Anggraini, "Analysis of Sequential Book Loan Data Pattern Using Generalized Sequential Pattern (GSP) Algorithm," *Int. J. Informatics Inf. Syst.*, vol. 2, no. 1, pp. 17–23, 2019, doi: 10.47738/ijiis.v2i1.10.

[8] Y. Feng, H. Chen, and L. He, "Consumer responses to femvertising: A data-mining case of Dove's 'Campaign for Real Beauty' on YouTube," *J. Advert.*, vol. 48, no. 3, pp. 292–301, 2019, doi: 10.1080/00913367.2019.1602858.

[9] N. Setiyawati, "Application of Association Rule Mining and Mining Sequential Patterns on Crm PT. Armada International Motor," *Creat. Commun. Innov. Technol. J.*, vol. 11, no. 1, pp. 95–101, 2018, doi: 10.33050/ccit.v11i1.562.

[10] E. A. Z. Noughabi, A. Albadvi, and B. H. Far, "How Can We Explore Patterns of Customer Segments' Structural Changes? A Sequential Rule Mining Approach," in *2015 IEEE International Conference on Information Reuse and Integration*, 2015, pp. 273–280, doi: 10.1109/IRI.2015.52.

[11] B. C. Kachhadiya and B. Patel, "A Survey on Sequential Pattern Mining Algorithm for Web Log Pattern Data," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, May 2018, pp. 1269–1273, doi: 10.1109/ICOEI.2018.8553691.

[12] D. S. Maylawati, H. Aulawi, and M. A. Ramdhani, "The concept of sequential pattern mining for text," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, p. 012042, Dec. 2018, doi: 10.1088/1757-899X/434/1/012042.

[13] M. Zaki, "Fast mining of sequential patterns in very large databases," 1997.

[14] K. Shudo, "Message bundling on structured overlays," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, Jul. 2017, pp. 424–431, doi: 10.1109/ISCC.2017.8024566.

[15] S. Rezig, Z. Achour, and N. Rezg, "Using Data Mining Methods for Predicting Sequential Maintenance Activities," *Appl. Sci.*, vol. 8, no. 11, p. 2184, Nov. 2018, doi: 10.3390/app8112184.

[16] W. Gan, J. C.-W. Lin, P. Fournier-Viger, H.-C. Chao, and P. S. Yu, "A Survey of Parallel Sequential Pattern Mining," *ACM Trans. Knowl. Discov. Data*, vol. 13, no. 3, pp. 1–34, Jun. 2019, doi: 10.1145/3314107.

[17] R. L. Helmreich and H. C. Foushee, "Why CRM? Empirical and Theoretical Bases of Human Factors Training," in *Crew Resource Management*, Elsevier, 2019, pp. 3–52, doi: 10.1016/B978-0-12-812995-1.00001-4.

[18] T. Guyet and R. Quiniou, "NegPSpan: efficient extraction of negative sequential patterns with embedding constraints," *Data Min. Knowl. Discov.*, vol. 34, no. 2, pp. 563–609, Mar. 2020, doi: 10.1007/s10618-019-00672-w.

[19] C. Pypno and G. Sierpiński, "Automated large capacity multi-story garage—Concept and modeling of client service processes," *Autom. Constr.*, vol. 81, pp. 422–433, Sep. 2017, doi: 10.1016/j.autcon.2017.03.006.

[20] M. Taufik, F. Renaldi, and F. R. Umbara, "Implementing Online Analytical Processing in Hotel Customer Relationship Management," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1115, no. 1, p. 012040, Mar. 2021, doi: 10.1088/1757-899X/1115/1/012040.

[21] J. Sterne, *Customer service on the Internet: building relationships, increasing loyalty, and staying competitive*. John Wiley \& Sons, Inc., 2000.