Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 6, December 2022, pp. 3422~3432
ISSN: 2302-9285, DOI: 10.11591/eei.v11i6.4021

A hybrid recommender system based on customer behavior and transaction data using generalized sequential pattern algorithm

Ramos Somya[1,2], Edi Winarko[1], Sigit Priyanta[1]
[1]Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2]Department of Informatic Engineering, Universitas Kristen Satya Wacana, Salatiga, Indonesia

Article Info

Article history:

Received Apr 30, 2022
Revised Jul 27, 2022
Accepted Aug 23, 2022

Keywords:

Customer behavior data
Generalized sequential pattern
Hybrid recommender system
Sequential pattern mining
Simple additive weighting

ABSTRACT

In the future, the quality of product suggestions in online retailers will influence client purchasing decisions. Unqualified product suggestions can result in two sorts of errors: false negatives and false positives. Customers may not return to the online store as a result of this. By merging sales transaction data and consumer behavior data in clickstream data format, this work offers a hybrid recommender system in an online store utilizing sequential pattern mining (SPM). Based on the clickstream data components, the product data whose status is only observed by consumers is assessed using the simple additive weighting (SAW) approach. Products with the two highest-ranking values are then coupled with product data that has been purchased and examined in the SPM using the generalized sequential pattern (GSP) method. The GSP algorithm produces rules in a sequence pattern, which are then utilized to construct product suggestions. According to the test results, product suggestions derived from a mix of sales transaction data and consumer behavior data outperform product recommendations generated just from sales transaction data. Precision, recall, and F-measure metrics values rose by 185.46, 170.83, and 178.43%, respectively.

Corresponding Author:

Edi Winarko
Department of Computer Science and Electronics, Universitas Gadjah Mada
North Sekip, BLS 21 Yogyakarta 55281, Indonesia
Email: ewinarko@ugm.ac.id

1. INTRODUCTION

Market basket analysis (MBA) is a strategy that can be used to discover relationships between items purchased by customers [1]. MBA is one of the methods in data mining that focuses on how to find purchasing patterns by extracting sales transaction data. One of the outcomes of the MBA process is the association rules of products purchased by customers [2]. The association rule discovers frequent itemsets of the purchased products in the database without considering the transaction orders. Sequential pattern mining (SPM), on the other hand, can be used to find patterns in the order in which products were bought in a database of transaction orders from customers [3].

One of the benefits of SPM in an online store is the ability to provide product recommendations to customers. Customers can buy several products at once in one transaction based on the product recommendations provided, so that it will be more efficient in the process of shipping goods from both the customer and seller sides [4]. The study of SPM for recommendation systems was conducted by Gunawan, which uses PrefixSpan to generate sequential patterns from an e-commerce dataset. This research shows that the SPM can produce a higher quality pattern for a recommendation system compared to association rule mining [5]. However, the online store recommendation system still requires further development to produce more accurate recommendations. One of the ideas to produce a better recommendation system is by combining sales transaction data with customer behavior data [6]–[8].

Customer behavior data provides a variety of important insights that can be used to understand shopping patterns and customer behavior before making a purchase or buying a product in an online store or other related applications [9], [10]. For example, what products are viewed, what products are searched for, what products are added to the shopping cart, and what products are finally purchased? [11] Customer behavior data in an online store is generated in the form of clickstream data. If the data is analyzed, it will have the potential to produce more accurate product recommendations for all products accessed rather than only analyzing sales transaction data [12], [13].

Recommendations that are not qualified can cause two types of errors, namely false negative (FN) and false positive (FP). FN is a list of products that are not recommended, even though customers like the product. While the FP is a list of recommended products, customers do not like the product. The type of error that must be avoided in online stores is a FP, because this error can cause customers to not repurchase at the online store [14].

Determining the novel recommendation system is challenging because it must be able to utilize data from various sources [15]–[17]. There are several previous studies on recommendation systems that have been developed. Lin and Jingtao [18] proposed a new idea in an online store recommendation system by involving contextual data such as the number of clicks on a product and product sales transaction data to generate product recommendations. The calculation of preference degree is carried out on the two data components using the arc tangent, so that the product item with the largest preference degree value is the most recommended product to consumers. This research hasn't been put into practice directly in online stores. Instead, it has been simulated using random sample data in calculations.

On the basis of social commerce, a novel model of tourism recommender system was developed [19]. The purpose of this research is to provide a recommendation system for tourist destinations by utilizing contextual data from customers. Collaborative filtering is used to analyze social media users based on their personal preferences, interests, and relationships. Based on experimental evidence, the recommendation system can generate recommended products and services in social commerce better than other common methods.

The K-means recommendation system [20] was developed by utilizing customer personal data such as age and gender to generate clustered customer profiles using the K-means method [21]. Each cluster where the customers live is analyzed using collaborative filtering to generate movie recommendations that fit into each cluster. Based on the test results, the proposed model can improve the quality of movie recommendations with its accuracy and performance [22], [23]. According to Liao *et al*. [24], a recommendation system for social media was developed by utilizing social media users' behavior data using clustering and an association rule approach. Data on the behavior of social media users was obtained using a questionnaire survey method. This study uses the clustering method to cluster the users into their most suitable groups based on the similarity of their profiles. The association rule method is then applied to each cluster to uncover the relationship between the purchased items.

Recommender systems offer products or services according to the users' preferences [25] by utilizing common data such as ratings, reviews, and feedback [26]–[28] to generate personalized recommendations [29], [30]. Recommender systems can be classified into several types based on the data used to generate recommendations. The hybrid recommender system utilizes information from user data and product data items (content-based filtering). In addition, the hybrid recommender system also uses information related to a set of users and their relationships to product items (collaborative filtering) [31], [32]. In other words, the hybrid recommender system is a combination of content-based filtering and collaborative filtering.

Based on the described background, in this study, a new hybrid recommender system was developed for online stores using the SPM approach. The novelty of this research is that it utilizes customer behavior data in the SPM to generate sequential patterns, since the common SPM utilizes purchased product data alone. Customer behavior data in the form of clickstream data is multi-criteria decision making (MCDM) data. Multi-criteria data is thought to produce more accurate predictions than single-criteria data [33]–[35]. In this research, the product data whose status is viewed only by the customers is analyzed using the simple additive weighting (SAW) method based on the clickstream data components. Product ranking results will be combined with purchased product data for further processing in the SPM using the generalized sequentials pattern (GSP) algorithm. The result of the SPM process with the GSP algorithm is sequential patterns that can be used to develop product recommendations. It is hoped that by adding customer behavior data in the form of clickstream data, product recommendations in online stores will be able to get better.

2. THE PROPOSED METHOD

The hybrid recommender system for online store developed in this study utilizes sales transaction data and customer behavior data in the form of clickstream data. In general, the description of this research is shown in Figure 1. Based on Figure 1, this research consists of 2 main stages, namely data gathering and data analysis. Sales transaction data and clickstream data will be collected from online store. After all of the data is stored in the database, the data analysis stage will be carried out. The first data to be analyzed is a list of products whose status is only seen by the customers. SAW method is used to rank this data based on the clickstream data components. The result of the ranking stage is product ranking data that will be selected in two highest-ranking values. The selected products will be combined with purchased product data from sales transaction data. The combined data is then analyzed in SPM using GSP algorithm. The outcomes of the SPM with the GSP algorithm are rules in sequence pattern, where these rules will be used to compile product recommendations and product bundling for online stores. Based on the block diagram, there are five main processes carried out in this research, namely:

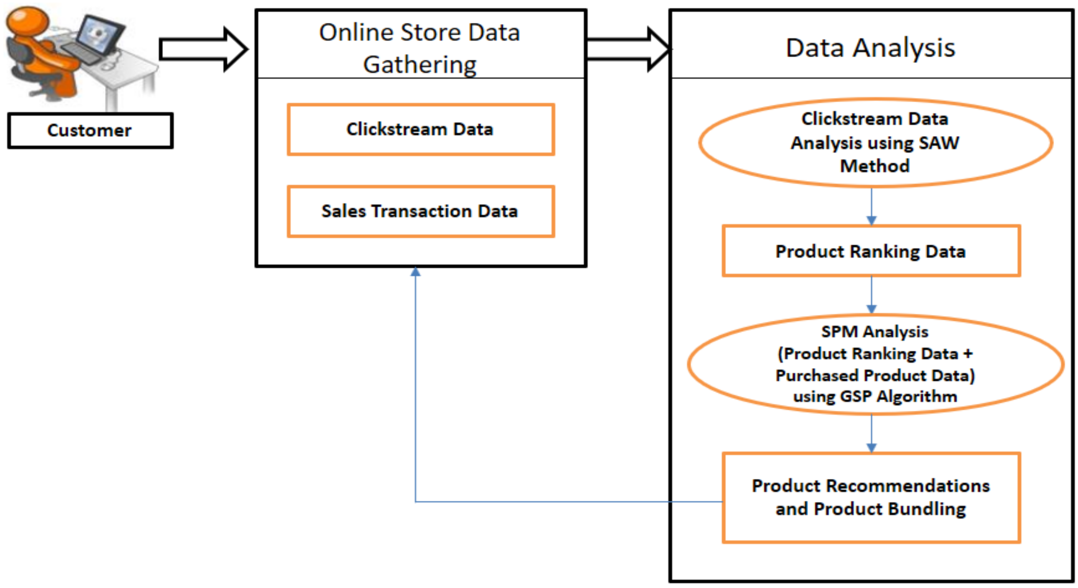

Figure 1. System block diagram

2.1. Developing an online store system

This study develops an online store website that will be used to record purchased item data and customer behavior data in clickstream data format. The products presented in the online store are specific to university's student daily needs, such as food, beverages, office equipment, and toiletries. The customer behavior data focuses on how customers decide to spend their data resources (time, money, and effort) on a product or service provided [36]. The 8 components of customer behavior data to be recorded are [37]:

- Product viewing time: how many seconds a product is seen by a customer in single purchase transaction.
- Number of product views: how many times a product is seen by a customer in single purchase transaction.
- Number of product searches: how many times a product is searched by a customer in the product search feature in single purchase transaction.
- History of a purchased product: history of how many times a product has been purchased.
- History of a product's viewing times: history of how many seconds a product is viewed by customers as long as it presented in the online store page.
- History of a product views: history of how many times a product is viewed by customers as long as it presented in the online store page.
- History of a product searches: history of how many times a product is searched by customers as long as it presented in the online store page.
- Product discount: the amount of discount on a product.

2.2. Dataset collection

The online store application that has been developed is operated to serve online shopping activities for customers. In this process, sales transaction data, and customer behavior data are recorded by the online store application and stored in a database.

2.3. Data analysis

The product data whose status is viewed only by the customers is analized using SAW method to determine which products can be selected to be combined with the list of products whose status is purchased. This method is used to determine the best alternative from a set of existing alternatives [38]. The stages of the SAW method are such as [39]:

a. Defining criteria (C) and preference weights (W).
b. Compiling decision matrix based on the critera (C) and normalizing the matrix according to the type of the attribute. Use (1) if the criterion is benefit attribute or use (2) if the criterion is cost attribute.

$$R_{ij} = \frac{X_{ij}}{\max X_{ij}} \quad (1)$$

$$R_{ij} = \frac{\min X_{ij}}{X_{ij}} \quad (2)$$

R_{ij} is the normalized performance rating, X_{ij} is the attribute value of each criterion, max X_{ij} is the greatest value of each criterion, and min X_{ij} is the smallest value of each criterion.

c. Calculating preference value for each alternative (V_i) using (3).

$$V_i = \sum_{j=1}^{n} W_j * R_{ij} \quad (3)$$

V_i is the ranking value for each alternative (A_i), W_j is the weighted value of each criterion, and R_ij is the normalized performance rating value.

d. Determining rank, the greater value of V_i will indicate the alternative A_i is preferred.

In this study, the alternative item (A_i) that will be ranked is derifed from a list of products whose status is only seen by the customers. The 8 components of the recorded clickstream data are used as criteria (C) in the ranking process. Decision maker determines the preference weight for each criterion. The total weight is 100%, so each criterion has 12.5% of weight. Each criterion is benefit attribute because in this case the greatest value is the best. Table 1 is a table of criteria and their preference weights.

Table 1. Criteria and their preference weights [37]

Code	Criteria (C_i)	Weight (W%)	Benefit or cost
C_1	Product viewing time	12.5	Benefit
C_2	Number of product views	12.5	Benefit
C_3	Number of product searches	12.5	Benefit
C_4	History of a purchased product	12.5	Benefit
C_5	History of a product's viewing times	12.5	Benefit
C_6	History of a product views	12.5	Benefit
C_7	History of a product searches	12.5	Benefit
C_8	Product discount	12.5	Benefit

The GSP algorithm is used for datasets that have a sequence, usually a sequence of transactions that occur within a certain time [40]. Table 2 shows a sequence dataset consists of purchased products and viewed only products from customers. Viewed only products from customer ID 1 in the first transaction are {abce}. For instance, the two highest-ranking products from the result of SAW method are {ac}. Therefore, the combination of products whose status is purchased and products with two highest-ranking values is {dghac}. This data combination will be analized in the SPM using GSP algorithm. GSP algorithm will extract this dataset to find sequential patterns [41]. The process of the GSP algorithm can be seen in Figure 2.

Table 2. Sequence dataset

Customer ID	Transaction time	Purchased product	Viewed only product
1	10, 20, 25	<{dgh}, {bf}, {agh}>	<{abce}, {cgh}, {bcde}>
2	10, 20	<{abf}>, {fgh}>	<{cdg}>, {abcd}>
3	15, 20	<{abf}, {e}>	<{cdeg}, {afgh}>
4	10, 15, 20, 25	<{cd}, {abc}, {abf}, {ac}>	<{abfg}, {dgh}, {cde}, {beg}>

There are two main steps in GSP algorithm, namely candidate generation and support counting [42]. Candidate generation stage has two steps, namely join phase and prune phase. Candidate sequences in join phase are generated by joining or merging frequent itemset (F_{k-1}) with itself. The set of candidates generated in this join phase will be denoted in candidate sequence (C_k). Prune phase removes candidate sequences that do not meet the specified minimum support value. All candidates who have a support value greater than or equal to a predetermined minimum support value are called frequent, which meet the requirements to be F_k. Support counting stage aims to find all candidates in a sequence dataset (*D*).

```
F1 = the set of frequent 1-sequence
k = 2,
do while F(k-1) is not null
        Generate candidate sets Ck (set of candidate k-sequences);
        For all input sequences s in the sequence dataset D
        do
                Increment count of all a in Ck if s supports a
        End do
        Fk = {a € Ck such that its frequency exceeds the threshold}
        k = k+1;
End do
Result = Set of all frequent sequences is the union of all fk's
```

Figure 2. GSP algorithm

2.4. Compiling product recommendation and product bundling

GSP algorithm generates sequence patterns in different combinations of product (L_1, L_2, … L_n). L_1 is sequence pattern consists of 1 product, L_2 is sequence pattern consists of 2 products, and so on. In this study, the sequence pattern chosen to be used as a product recommendation and product bundling is the sequence pattern with the 2 highest values in each combination. The 2 highest values are based on the support count, support, and confidence values in each generated sequential pattern combination. For example, in L_2 there are 5 sequence patterns with support count, support, and confidence values of 5.00, 0.75, and 0.67, respectively. Then there are 10 sequence patterns with values of 4.00, 0.65, and 0.50, and there are 20 sequence patterns with values of 4.00, 0.33, and 0.25. Then the sequence patterns chosen as the product recommendation is the sequence patterns with the 2 highest values for count, support, and confidence values, namely 5 sequence patterns with values of 5.00, 0.75, and 0.67, and 10 sequence patterns with values of 4.00, 0.65, and 0.50.

The selected sequence patterns will be used to compile product recommendation and product bundling for online store. Each product in the online store will be given 4 recommended products. Meanwhile, the product bundling is composed by pairing 2 different products according to the selected sequence patterns. The product bundling is then presented in the special page in the online store.

2.5. Testing the product recommendation and product bundling

The testing phase of the results of data analysis will be carried out by comparing the rules generated from the SPM process using only sales transaction data and the rules generated from the SPM process, which combines sales transaction data with customer behavior data. The results of the rules from the two processes will be implemented to provide product recommendations and product bundling to online store customers. Testing the quality of product recommendations and product bundling is carried out by calculating the precision, recall, F-measure, precision @K, recall @K, and F-measure @K values and the results will be compared. The calculation of the value of precision, recall, and F-measure is shown in (4)-(6) [43].

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F\text{-}Measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

True positive (TP) on information retrieval is positive data that is detected correctly, while FP is negative data but detected as positive data. FN is the opposite of TP, where data is positive, but is detected as negative data. The values of precision @K, recall @K, and F-measure @K will be determined using the varible of K, which is the rank order derifed from the values of support count, support, and confidence of the product recommendations. This calculation will be explained in details in the results and discussion section.

3. RESULTS AND DISCUSSION

In this section, it is explained the results of the research, and at the same time is given the comprehensive discussion. The discussion is be made in several sub-sections.

3.1. Result of data analysis

After operating for one month, the online store managed to collect 102 sales transaction data from 33 customers. This sales transaction data consists of 206 different products with purchased status and only viewed. In addition, eight clickstream data components were also successfully recorded in the database. The calculation of the SAW method is carried out by the online store website on every sales transaction data, which is specifically for product data with only viewing status. The result of the calculation from SAW is in the form of a product ranking order, so that it can be seen what products have the potential to be considered as recommendations. The example of the recorded sales transaction data, clickstream data and product ranking results is shown in Table 3.

Table 3. Sales transaction data, clickstream data, and ranking results

Product name	Quantity	Price (Rp.)	Sub total	Status	Criteria								V
					C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8 (Rp.)	
Rexona deodorant free spirit 50 Ml	0	18.200,00	0	Only seen	30	1	1	1	120	3	3	300,00	1.00
Protecal vitamin C a calsium 10'S orange	0	41.400,00	0	Only seen	30	1	1	1	70	3	3	800,00	0.90
Red bull energy drink 250 Ml	0	19.800,00	0	Only seen	10	1	1	1	30	3	3	400,00	0.75
Enervon-C vitamin 4'S tablet	0	5.900,00	0	Only seen	10	1	1	2	50	3	3	300,00	0.75
Luwak white coffee less sugar 10X20g	4	13.500,00	Rp. 52.800,00	Purchased	30	1	1	5	70	3	3	300,00	OK

Table 3 is one of the sales transaction data that has been successfully recorded in the database. In this data, the customer buys one product, and there are four products whose status is viewed only by the customer. Product with "purchased" status is labeled OK, while products whose status is "only seen" are displayed using a ranking, where the value of V is determined based on the calculation of the SAW method.

Based on the results of ranking on product data whose status is only viewed, the products with two highest-ranking values in each sales transaction data will be selected and combined with products whose status is purchased. All product data will then be analyzed in the SPM using the GSP algorithm in Python. The first and the second dataset come from the same transaction period, where there were 102 sales transactions from 33 customers. The first dataset has 321 records and consists of purchased product alone, while the second dataset has 452 records, consists of purchased products plus a list of products with viewing only status with the 2 highest ranking values. The example of the dataset can be seen in Table 4 and Table 5.

Table 4. Example of dataset 1: consists of purchased products only

Customer ID	Order ID	Date	Product name	Status	V
672019114	267	19/04/2021 09:35	Kiky Double Line Paper/10	Purchased	OK
672019114	267	19/04/2021 09:35	Wrigley's Candy Gum Doublemint 15G	Purchased	OK
672019114	267	19/04/2021 09:35	Indomie Fried Instant Noodles Plus Special 85G	Purchased	OK
672019114	267	19/04/2021 09:35	Samyang Fried Chicken Instant Noodles 130G	Purchased	OK

Table 5. Example of dataset 2: consists of purchased and viewed products with the 2 highest rankings

Customer ID	Order ID	Date	Product name	Status	V
672019255	285	20/04/2021 06:57	Delfi Chocolate Cashew 27G	Purchased	OK
672019255	285	20/04/2021 06:57	Frisian Flag Chocolate Milk 560G	Purchased	OK
672019255	285	20/04/2021 06:57	Indomie Fried Instant Noodles Plus Special 85G	Purchased	OK
672019255	285	20/04/2021 06:57	Abc Sardines Chili 155G	Only Seen	1.000
672019255	285	20/04/2021 06:57	Betadine Solution 30Ml	Only Seen	1.000

Columns used for sequence pattern search are customer ID, order ID, and product name columns. The sequence pattern search for the first dataset is carried out with a minimum support value of 2 or 0.06. The sequence patterns that have been successfully generated are 138 rules consisting of 73 rules with a combination of 1 product (L_1), 56 rules with a combination of 2 products (L_2), 8 rules with a combination of 3 products (L_3), and 1 rule with a combination of 4 products (L_4). One of the generated rules is *<{'Axe Deodorant Bodyspray Harumkan Indonesia 135 M', 'Beng-Beng Wafer Chocolate 20 G'}{'Cip Corned Beef 198 G'}>*. The rule consists of 3 product combinations (L_3), in which there are two sequential patterns, namely if the customers buy *Axe Deodorant Bodyspray Harumkan Indonesia 135 M* product and *Beng-Beng Wafer Chocolate 20 G* product in the first transaction, then in the second transaction, they will buy *Cip Corned Beef 198 G* product. This rule is supported by a support count value of 2, a support value of 0.061, and a confidence value of 0.500.

Sequence pattern search for the second dataset is also carried out with a minimum support value of 2 or 0.06. The sequence patterns that were successfully generated from the second dataset were 442 rules. This rule is a sequential pattern consisting of 86 rules on L_1, 232 rules on L_2, 98 rules on L_3, 19 rules on L_4, 6 rules on L_5, and 1 rule on L_6. The number of rules generated from the second dataset is more than the first dataset, because the first dataset only consists of a list of products with purchased status, while the second dataset consists of a list of products with purchased status plus products whose status is viewed only with the two highest ranking values.

3.2. Determining product recommendation and product bundling

Based on the sequence patterns generated with the GSP algorithm, the product recommendations and product bundling are then compiled. Each product will be given 4 product recommendations. In addition, the sequence patterns also used to compile product bundling. Product bundling contains 2 products that have a relationship based on the sequence patterns. Table 6 and Table 7 are several examples of the product recommendation and product bundling.

Table 6. Two examples of the product recommendation

List of product recommendations		Support count
Product name	Recommended product	
Cadbury Chocolate Dairy Milk 30 G	Cip Corned Beef 198 G	2
	Choki Choki Chocolate 4×10 g	0
	Delfi Chocolate Wafer Take-It 4 Fingers 35 G	0
	Beng-Beng Wafer Chocolate 20 G	0
Cip Corned Beef 198 G	Cadbury Chocolate Dairy Milk 30 G	2
	Axe Deodorant Bodyspray Harumkan Indonesia 135 M	2
	Beng-Beng Wafer Chocolate 20 G	2
	Ayam Brand Tuna Chunks in Water 185 g	0

Table 7. Two examples of product bundling

Number	Product bundling	Support count	Total support
1	Cadbury Chocolate Dairy Milk 30G	2	10
	Cip Corned Beef 198G	8	
2	Choki Choki Chocolate 4X10g	2	7
	Fresh Care Ointment Aroma Therapy 10Ml	5	

Table 6 is the example of product recommendations arrangement based on the sequence patterns. The *Cadbury Chocolate Dairy Milk 30 G* product only has a sequence pattern with the *Cip Corned Beef 198 G* product (the support count value is 2), so the three other products will be selected according to the similarity in the product category, namely *Choki Choki Chocolate 4X10g*, *Delfi Chocolate Wafer Take-It 4 Fingers 35 G*, and *Beng-Beng Wafer Chocolate 20 G*, so that there are 4 recommended products for each product in online store page display. Product bundlings in Table 7 are compiled based on the sequence patterns. *Cadbury Chocolate Dairy Milk 30 G* product and *Cip Corned Beef 198 G* product are sequentially related with total support value of 10, so these two products can be arranged in one bundling product.

3.3. Product recommendation rule testing

The product recommendations and product bundling have been applied in online stores for 45 days, from July to early September 2021. Customers are given four product recommendations for each product accessed. Meanwhile, product recommendations in the form of product bundlings are presented on a special menu on the online store page display. The same testing mechanism was also carried out on the rules generated from the second dataset, where product recommendations were applied in online stores for 45 days, from early September to mid-October 2021.

Based on two periods of sales transactions that have been carried out, sales transaction data was collected. The transaction data is then calculated to obtain the precision, recall, and F-measure values, where these values are determined based on the (*TP*), (*FP*), and (*FN*) values. Calculations for testing the application of product recommendations and product bundling are also carried out using the precision @K, recall @K, and F-measure @K values, which are based on the ranking of support count values, support values, and confidence values for each recommended product and product bundling. Table 8 shows some of the purchase transaction data from the database.

Table 8. Purchase transaction data

Customer ID	Viewed product	Recommended product	Purchased product
672019069	Arnott's Chocolate Tim Tam Chocolate 81 G	Dove Shampoo Hair Fall Treatment 160 Ml	Arnott's Chocolate Tim Tam Chocolate 81 G
	Close Up Toothpaste Gel Green Menthol Fresh 65 G	Axe Deodorant Bodyspray Harumkan Indonesia 135 M	Close Up Toothpaste Gel Green Menthol Fresh 65 G
	Dove Shampoo Hair Fall Treatment 160 Ml	Fresh Care Ointment Aroma Therapy 10 Ml	Dove Shampoo Hair Fall Treatment 160 Ml
	Wardah Seaweed Balancing Facial Scrub 60 g	Antangin Jrg Catch a Cold Medicine Syrup 5×15 ml	Wardah Seaweed Balancing Facial Scrub 60 g
		Close Up Toothpaste Gel Green Menthol Fresh 65 G	

A customer in Table 8 accessed four products in the online store. At the same time, four product recommendations are also accessed. In this test scenario, only product recommendations from sequential patterns are listed in the calculation. The same product recommendations from viewed products will be recorded once. Based on Table 8, the number of product recommendations purchased by consumers (*TP*) is 2 (*Close Up Toothpaste Gel Green Menthol Fresh 65 G* product and *Dove Shampoo Hair Fall Treatment 160 Ml* product). The number of product recommendations that are not purchased by consumers (*FP*) is 3 (*Axe Deodorant Bodyspray Harumkan Indonesia 135 M* product, *Fresh Care Ointment Aroma Therapy 10 Ml* product, and *Antangin Jrg Catch a Cold Medicine Syrup 5×15 ml* product). The number of products that are not recommended but are purchased by consumers (*FN*) is 2 (*Arnott's Chocolate Tim Tam Chocolate 81 G* product, and *Wardah Seaweed Balancing Facial Scrub 60 g* product). So, the values of precision, recall, and F-measure according to (4)-(6) are 0.400, 0.500, and 0.222. Calculations for testing the application of product recommendation and product bundling are also carried out using the precision @K, recall @K, and F-measure @K values, which are based on the ranking of support count values, support values, and confidence values for each product recommendation and product bundling. One of the calculation results is shown in Table 9.

Table 9. The calculation of Precision @K and Recall @K

Customer ID	Product recommendations	K	Precision @K	Recall @K
672019069	Dove Shampoo Hair Fall Treatment 160Ml (purchased)	1	1 / 1 = 1.000	1 / 5 = 0.200
	Close Up Toothpaste Gel Green Menthol Fresh 65G (purchased)	2	2 / 2 = 1.000	2 / 5 = 0.400
	Fresh Care Ointment Aroma Therapy 10Ml (only seen)	3	2 / 3 = 0.667	2 / 5 = 0.400
	Antangin Jrg Catch a Cold Medicine Syrup 5X15ml (only seen)	4	2 / 4 = 0.500	2 / 5 = 0.400
	Axe Deodorant Bodyspray Harumkan Indonesia 135M (only seen)	5	2 / 5 = 0.400	2 / 5 = 0.400
	Average value		0.713	0.360

Table 9 is the calculation of precision@K and recall @K. The values are determined using the ranking order of *K*, which is derifed from support count values, support values, and confidence values for each product recommendation. Based on the calculation, precision@K and recall @K values of the product recommendations are 0.713 and 0.360. Meanwhile, F-measure@K value is 0.478. The summary of the whole test results is shown in Table 10.

The results of testing the application of product recommendations on online stores are: the precision value in the first scenario is 11.00%, while in the second scenario it is 31.40% (increased by 185.46%). The recall value in the first scenario is 9.60%, while in the second scenario it is 26.00% (increased by 170.83%). The F-measure value in the first scenario is 10.20%, while in the second scenario it is 28.40% (increased by 178.43%). The results of testing the application of product bundling recommendations on online stores are that the values of precision, recall, and F-measure in the first scenario are 20.00%, 25.00%, and 22.20%. In the second scenario, the test results were 60% (an increase of 200%), 30% (an increase of 20%), and 40% (an increase of 80.18%).

The results of testing the application of product recommendations on online stores by considering the order of recommendation rules are: precision @K, recall @K, and F-measure @K values in the first scenario are 11.00%, 6.00%, and 7.80%. In the second case, the metric values are 29.80% (up by 170.91%), 16.70% (up by 178.33%), and 21.40% (up by 174.36%). The results of testing the application of product bundling recommendations on online stores by considering the ranking order of the recommendation rules are: precision @K, recall @K, and F-measure @K values in the test in the first scenario are 35.40%, 15.00%, and 21.10%. In the second case, the metric values are 65.10% (an increase of 83.90%), 33.30% (an increase of 122.00%), and 44.10% (an increase of 109.01%).

Table 10. Recommendation rule test results

The implementation of sequence pattern for recommendation in online store	Metrics	Scenario: Recommendation using sales transaction data	Scenario: Recommendation using sales transaction data + clickstream data
Product recommendation	Precision	0.110	0.314
	Recall	0.096	0.260
	F-Measure	0.102	0.284
Product bundling	Precision	0.200	0.600
	Recall	0.250	0.300
	F-Measure	0.222	0.400
Product recommendation based on the order of rank values results	Precision @K	0.110	0.298
	Recall @K	0.060	0.167
	F-Measure @K	0.078	0.214
Product bundling based on the order of rank values results	Precision @K	0.354	0.651
	Recall @K	0.150	0.333
	F-Measure @K	0.211	0.441

4. CONCLUSION

Customer clickstream data and sales transaction data have shown to be superior than product suggestions based only on sales transaction data, according to the findings of the test. Precision, recall, and F-measure values increased by 185.46%, 170.83%, and 178.43% when the suggested goods were used. The use of product bundling suggestions raised the precision, recall, and F-measure values by 200.00%, 20.00%, and 80.18%. Precision @K, recall @K, and F-measure @K rose by 170.91%, 178.33%, and 174.36% in the application of product recommendations. To sum it up, the execution of product bundling suggestions enhanced the values of precision @K, recall @K, and F-measure @K by 83.90%, 122.00%, and 109.01%. This study also shows that the output of the SPM may be improved by merging sales transaction data with customer behavior data.

Other clickstream data components generated by online shop websites and mobile commerce applications must be examined in future studies. Customer contextual data such as location data and customer background may all be shown, for example in a shopping path component. More clickstream data components are expected to lead to more accurate product suggestions in an online store by using them. The testing of product suggestions in an online store may be included into several sales transactions. Precision, recall, and F-measure all stand to benefit from increased sales of items that were suggested to customers.

ACKNOWLEDGEMENTS

This work is supported by Universitas Gadjah Mada and Universitas Kristen Satya Wacana. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] A. B. Rao, J. S. Kiran, and G. Poornalatha, "Application of market–basket analysis on healthcare," *International Journal of System Assurance Engineering and Management*, Aug. 2021, pp. 1-6, doi: 10.1007/s13198-021-01298-2.
[2] A. A. Aldino, E. D. Pratiwi, Setiawansyah, S. Sintaro, and A. D. Putra, "Comparison of market basket analysis to determine consumer purchasing patterns using fp-growth and apriori algorithm," *2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, 2021, pp. 29-34, doi: 10.1109/ICOMITEE53461.2021.9650317.
[3] C. Le, K. J. Shrestha, H. D. Jeong, and I. Damnjanovic, "A sequential pattern mining driven framework for developing construction logic knowledge bases," *Automation in Construction*, vol. 121, p. 103439, 2021, doi: 10.1016/j.autcon.2020.103439.

[4] Y. Fang, X. Xiao, X. Wang, and H. Lan, "Customized bundle recommendation by association rules of product categories for online supermarkets," in *IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Jul. 2018, pp. 472–475, doi: 10.1109/DSC.2018.00076.
[5] R. Gunawan, "Online retail pattern quality improvement: from frequent sequential pattern to high-utility sequential pattern," in *4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Dec. 2021, pp. 242-246, doi: 10.1109/ISRITI54043.2021.9702782.
[6] K. Tatiana, and M. Mikhail, "Market basket analysis of heterogeneous data sources for recommendation system improvement," *Procedia Computer Science*, vol. 136, pp. 246-254, 2018, doi: 10.1016/j.procs.2018.08.263.
[7] L. C. Annie, and A. D. Kumar, "Market basket analysis for a supermarket based on frequent itemset mining," *International Journal of Computer Science Issues*, vol. 9, no. 5, pp. 257–264, Sep. 2012.
[8] J. Iwanaga, N. Nishimura, N. Sukegawa, and Y. Takano, "Improving collaborative filtering recommendations by estimating user preferences from clickstream data," *Electronic Commerce Research and Applications*, vol. 37, p. 100877, Sep.-Oct. 2019, doi: 10.1016/j.elerap.2019.100877.
[9] A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, A. Almaghthawi, and E. A. A. Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 539-549, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.
[10] G. Morshed, H. Ujir, and I. Hipiny, "Customer's spontaneous facial expression recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, pp. 1436-1445, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1436-1445.
[11] G. Lin, Y. Miao, and S. Liu, "Decision-behavior based online shopping," in *International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Dec. 2018, pp. 1321-1326, doi: 10.1109/ICARCV.2018.8581336.
[12] G. Pal, G. Li, and K. Atkinson, "Big data real-time clickstream data ingestion paradigm for e-commerce analytics," in *4th International Conference for Convergence in Technology (I2CT)*, Oct. 2018, pp. 1-5, doi: 10.1109/I2CT42659.2018.9058112.
[13] J. Yeo, S. -w. Hwang, s. kim, E. Koh, and N. Lipka, "Conversion prediction from clickstream: modeling market prediction and customer predictability," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 2, pp. 246-259, 1 Feb. 2020, doi: 10.1109/TKDE.2018.2884467.
[14] E. Mena-Maldonado, R. Cañamares, P. Castells, Y. Ren, and M. Sanderson, "Agreement and disagreement between true and false-positive metrics in recommender systems evaluation," in *SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2020, pp. 841-850, doi: 10.1145/3397271.3401096.
[15] A. Da'u, N. Salim, I. Rabiu, and A. Osman, "Weighted aspect-based opinion mining using deep learning for recommender system," *Expert Systems with Applications*, vol. 140, p. 112871, Feb. 2020, doi: 10.1016/j.eswa.2019.112871.
[16] T. Oktavia, and S. Sujarwo, "A meta-learning recommender system framework for identifying learning partner," *ICIC Express Letters*, vol. 14, no. 2, pp. 117-124, Feb. 2021, doi: 10.24507/icicel.15.02.117.
[17] S. Singhal, and P. Tanwar, "A prediction model for benefitting e-commerce through usage of regional data: A new framework," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 4, pp. 1009-1018, Dec. 2021, doi: 10.11591/ijai.v10.i4.pp1009-1018.
[18] D. Lin and S. Jingtao, "A recommender system based on contextual information of click and purchase data to items for e-commerce," *Third International Conference on Cyberspace Technology (CCT 2015)*, 2015, pp. 1-6, doi: 10.1049/cp.2015.0823.
[19] L. Esmaeili, S. Mardani, S. A. H. Golpayegani, and Z. Z. Madar, "A novel tourism recommender system in the context of social commerce," *Expert Systems with Applications*, vol. 149, p. 113301, Jul. 2020, doi: 10.1016/j.eswa.2020.113301.
[20] B. A. Jaafar, M. T. Gaata, and M. N. Jasim, "Home appliances recommendation system based on weather information using combined modified k-means and elbow algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 3, pp. 1635-1642, Sep. 2020, doi: 10.11591/ijeecs.v19.i3.pp1635-1642.
[21] S. Al-Otaibi *et al*., "Cosine similarity-based algorithm for social networking recommendation," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 1881-1892, Apr. 2022, doi: 10.11591/ijece.v12i2.pp1881-1892.
[22] A. Yassine, L. Mohamed, and M. A. Achhab, "Intelligent recommender system based on unsupervised machine learning and demographic attributes," *Simulation Modelling Practice and Theory*, vol. 107, p. 102198, Feb. 2021, doi: 10.1016/j.simpat.2020.102198.
[23] A. Ez-Zahout, H. Gueddah, A. Nasry, R. Madani, and F. Omary, "A hybrid big data movies recommendation model based k-nearest neighbors and matrix factorization," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 434-441, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp434-441.
[24] S-H. Liao, R. Widowati, and Y-C. Hsieh, "Investigating online social media users' behaviors for social commerce recommendations," *Technology in Society*, vol. 66, p. 101655, Aug. 2021, doi: 10.1016/j.techsoc.2021.101655.
[25] B. Hssina, A. Grota, and M. Erritali, "Recommendation system using the K-nearest neighbors and singular value decomposition algorithms," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5541-5548, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5541-5548.
[26] N. N. Qomariyah, D. Kazakov, and A. N. Fajar, "On the benefit of logic-based approach to learn pairwise comparisons," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 6, pp. 2637-2649, Dec. 2020, doi: 10.11591/eei.v9i6.2384.
[27] P. P. Rokade, and A. Kumari, "Business recommendation based on collaborative filtering and feature engineering – aproposed approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9. no. 4, pp. 2614-2619, Aug. 2019, doi: 10.11591/ijece.v9i4.pp2614-2619.
[28] J. Kim, D. Hwang, and H. Jung, "Product recommendation system based user purchase criteria and product reviews," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 5454-5462, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5454-5462.
[29] A. H. Nasyuha *et al*., "Frequent pattern growth algorithm for maximizing display items," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 19, no. 2, pp. 390-396, Apr. 2021, doi: 10.12928/telkomnika.v19i2.16192.
[30] S. Babeetha, B. Muruganantham, S. G. Kumar, and A. Murugan, "An enhanced kernel weighted collaborative recommended system to alleviate sparsity," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 447-454, Feb. 2020, doi: 10.11591/ijece.v10i1.pp447-454.
[31] A. H. Espejel and F. J. Cantu-Ortiz, "Data mining techniques to build a recommender system," *2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC)*, 2021, pp. 217-221, doi: 10.1109/ISCSIC54682.2021.00047.
[32] N. Thongsri, P. Warintarawej, S. Chotkaew, and W. Saetang, "Implementation of a personalized food recommendation system based on collaborative filtering and knapsack method," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 1, pp. 630-638, Feb. 2022, doi: 10.11591/ijece.v12i1.pp630-638.
[33] E. Hikmawati, N. U. Maulidevi, and K. Surendro, "Adaptive rule: A novel framework for recommender system," *ICT Express*, vol. 6, no. 3, pp. 214-219, Sep. 2020, doi: 10.1016/j.icte.2020.06.001.

[34] Q. Shambour, "A deep learning based algorithm for multi-criteria recommender systems," *Knowledge-Based Systems*, vol. 211, p. 106545, Jan. 2021, doi: 10.1016/j.knosys.2020.106545.
[35] K. Zhang, X. Liu, W. Wang, and J. Li, "Multi-criteria recommender system based on social relationships and criteria preferences," *Expert Systems with Applications*, vol. 176, p. 114868, Aug. 2021, doi: 10.1016/j.eswa.2021.114868.
[36] J. George, "Growing & changing trends in consumer behavior," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 4804-4809, doi: 10.1109/ICEEOT.2016.7755633.
[37] R. Somya, E. Winarko, and S. Privanta, "A novel approach to collect and analyze market customer behavior data on online shop," *2021 2nd International Conference on Innovative and Creative Information Technology (ICITech)*, 2021, pp. 151-156, doi: 10.1109/ICITech50181.2021.9590161.
[38] P. C. Fishburn, *A Problem-based selection of multi-attribute decision making methods*, Hoboken, New Jersey, USA: Blackwell Publishing, 1967.
[39] K. R. MacCrimmon, *Decision making among Multiple-attribute alternatives: a survey and consolidated approach*, Santa Monica, California, USA: The Rand Corporation, 1968.
[40] M. Muhajir, and B. R. Efanna, "Association rule algorithm sequential pattern discovery using equivalent classes (SPADE) to Analyze the genesis pattern of landslides in Indonesia," *International Journal of Advances in Intelligent Informatics*, vol. 1, no. 3, pp. 158–163, Nov. 2015, doi: 10.26555/ijain.v1i3.50.
[41] M. J. Zaki, "Fast mining of sequential patterns in very large databases," *Technical Report*, Computer Science Department, The University of Rochester, New York, USA, 1997.
[42] R. Srikant, and R. Agrawal, "Mining sequential patterns: generalizations and performance improvements," in *Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology (EDBT '96)*, 1996, pp. 3-17, doi: 10.1007/BFb0014140.
[43] C. D. Manning, P. Raghavan, and H. Schütze, *An introduction to information retrieval*, Cambridge, England: Cambridge University Press, p. 482, 2008.

BIOGRAPHIES OF AUTHORS

Ramos Somya obtained his master degree in 2012 from Universitas Kristen Satya Wacana, Salatiga, Indonesia. He is a lecturer in Department of Informatic Engineering, Universitas Kristen Satya Wacana, Salatiga, Indonesia. Currently, he is a student of Doctorate Program of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interests are in business intelligence, data mining, machine learning, decision support system, and software engineering. He can be contacted at email: ramos.somya@mail.ugm.ac.id.

Edi Winarko obtained his Ph.D degree from Flinders University, Australia. He is a lecturer in Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interests are in data mining, machine learning, and information retrieval. He can be contacted at email: ewinarko@ugm.ac.id.

Sigit Priyanta obtained his Doctorate degree in 2016 from Doctoral Program in Computer Science, Universitas Gadjah Mada, Yogyakarta, Indonesia. He is a lecturer in Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. His research interests are in geographics information system, location-based services, and mobile information system. He can be contacted at email: seagatejogja@ugm.ac.id.