OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 7, Issue. 2, August 2022. pp. 53-64
doi:10.34818/indojc.2022.7.2.650

E-Commerce Recommender System on the Shopee Platform Using Apriori Algorithm

Rachmi Helfianur [#1], Z. K. A. Baizal[#2*]

 # School of Computing, Telkom University
Jl. Telekomunikasi No. 1, Bandung 40257, Indonesia

[1] rahmihelfianur@student.telkomuniversity.ac.id
[2] baizal@telkomuniversity.ac.id

Abstract

The development of E-Commerce continues to increase every year, and all online shopping platforms continue to increase competition. Shopee as one of the online shopping platforms offers various product categories that are needed by users. To make it easier for users when shopping online, it is necessary to implement a recommender system. Many previous researchers have developed a recommender system in e-commerce using a method based on content-based filtering, where this recommender system provides recommendations based on what users like and user activity, but does not provide optimal performance for the recommender system. Therefore, this research, we will develop a recommender system using the apriori algorithm. This research explores the scope of recommendations based on frequent item sets by applying an apriori algorithm to find items or products that are frequently purchased and frequently searched for. The apriori algorithm was selected because there are association rules that result in higher accuracy, and the results of the association rules will be used as a reference in determining the item set to be recommended. In this research, we take the domain of electronic goods in e-commerce. In this study, the system produces the highest and best association rules values with support values of 0.01, confidence 1.00, and lift 97.35 which can assist in making decisions in determining the itemset to be recommended and increasing user confidence in the recommendations provided and to avoid overspecialization.

Keywords: Recommender System, Apriori Algorithm, E-commerce, Association Rules

Abstrak

Perkembangan E-Commerce terus mengalami peningkatan setiap tahun, dan semua online shopping platform terus meningkatkan persaingan. Shopee sebagai salah satu online shopping platform menawarkan berbagai macam kategori produk yang dibutuhkan oleh para pengguna. Untuk mempermudah para pengguna saat berbelanja online, maka perlu diterapkan sebuah sistem rekomendasi. Penelitian sebelumnya telah banyak yang menggembangkan sistem pemberi rekomendasi pada sebuah platform e-commerce dengan berdasarkan content-based filtering, dimana system pemberi rekomendasi ini memberikan rekomendasi berdasarkan apa yang disukai oleh pengguna dan aktiviasi pengguna, namun kurang memberikan kinerja yang optimal terhadap sistem pemberi rekomendasi. Oleh karena itu, pada penelitian ini kami akan membangun system rekomendasi dengan menggunakan algoritma apriori. Penelitian ini mengeksplorasikan cakupan rekomendasi berdasarkan frequent item set dengan menerapkan algoritma apriori untuk menemukan item atau produk yang sering dibeli dan sering dicari. Algoritma apriori dipilih karena aturan asosiasi yang dihasilkan memiliki akurasi yang lebih tinggi, dan hasil dari aturan asosiasi algoritma apriori digunakan sebagai acuan dalam menentukan item set yang akan direkomendasikan. Dalam penelitian ini, kami mengambil domain barang eloktronik pada e-commerce. Penelitian ini, sistem menghasilkan nilai association rules tertinggi dan terbaik dengan nilai support 0.01, confidence 1.00, dan lift 97.35 yang dapat membantu dalam mengambil keputusan dalam menentukan itemset

*Corresponding Author

Received on April, 2022. Accepted on July, 2022

yang akan direkomendasikan dan meningkatkan kepercayaan pengguna terhadap rekomendasi yang diberikan, dan untuk menghindari overspecialization.

Kata Kunci: Sistem Rekomendasi, Algoritma Apriori, E-commerce, Association Rules

I. Introduction

THE developments in the world of e-commerce continue to increase every year so users experience difficulties and experience problems in choosing products that are safe, efficient, and user-friendly. One of the most popular online shopping platforms is shopee because shopee is an application that is easily accessible by users in conducting online shopping activities, shopee also offers a wide range of electronic products to products for daily needs [1]. Currently, many internet users are shopping online to meet their needs, therefore a recommender system is needed that makes it easy for users to choose and buy needs based on what they want. The purpose of the recommender system is to increase sales of a product and provide recommendations for the products that users are looking for, as well as be a consideration for customers in buying and choosing the products they want. The recommender system itself is a system that will produce a recommendation display for a product by using the methods contained in the recommender system.

The recommender system is a system used to get the desired product. Recommender systems are often used in an E-commerce system. Because using a recommender system can help users to offer the desired products and make it easier for users to make decisions when buying the desired product [2]. The purpose of the recommender system is to increase sales of a product. The recommender system itself consists of several methods, namely Collaborative-filtering, Content-based filtering, Knowledge-based, and Hybrid filtering [3]

Many previous studies have developed a recommender system in e-commerce by using the content-based filtering method, where this recommender system provides recommendations based on what users like and user activity, but does not provide optimal performance for the recommender system [4]. Therefore, this study uses the Apriori Algorithm to display the results of the recommendations, the Apriori algorithm was chosen because it has association rules that produce higher and best accuracy based on the support, confidence, and lift values. This study also explores the scope of recommendations based on frequent item sets by applying an apriori algorithm to find items or products that are frequently purchased and frequently searched. The association rules are obtained by the support and confidence of item relationships. The association rule has a high accuracy if the support value is greater than the minimum support and also if the confidence value is greater than the minimum confidence. Historical data of consumer purchases is used to form a combination of items. If the item combination meets the minimum support and confidence parameters, it can be used to determine the recommendations.

II. Literature Review

A. Recommender System

The recommender system is a system used by internet users to get the products they want. recommender system are often used in an e-commerce system. Because using a recommender system can help users to offer the desired products and make it easier for users to make decisions when buying the products they want [9]. The purpose of the recommender system is to increase sales of a product. The recommender system itself consists of several methods used, namely Collaborative-filtering, Content-based filtering, Knowledge-based, and Hybrid filtering [4].

a. ***Collaborative- filtering*****:** Providing recommendations based on feedback from users or consumers.
b. ***Content-based filtering*****:** Provide recommendations based on the similarity of the attributes of the item or item. For example, based on the type of goods, product materials, and others.
c. ***Knowledge-based*****:** Provide recommendations based on the condition of the attribute value specified by the user.

d. ***Hybrid filters*:** The combination of all methods to produce more accurate recommendations. Hybrid appears to overcome the weaknesses and limitations of a pure recommender system and a single recommender system.

B. Apriori Algorithm

Apriori algorithm is a method to find a pattern of relationship between one or more items in the dataset [5]. The apriori algorithm is also one of the data mining algorithms to search for frequent itemsets with Association rules. The Apriori algorithm uses a predetermined attribute frequency to be able to find out further information. In this apriori algorithm, candidate recommendations that may appear will be determined by taking into account the minimum support and minimum confidence values.

The apriori algorithm is one of the algorithms proposed by Agrawal and Srikant in 1994 which serves to determine frequent itemsets in Boolean associations [13]. This algorithm is in charge of monitoring the candidate development of the itemset [12]. The apriori algorithm uses a level-wise search approach, where k-itemset is used to get (k+1)-itemset.

C. Association Rules

Association Rule is a data mining technique used to find associative rules in a combination of items (Kusrini and Taufiq Luthfi, 2009) [13]. There are several parameters in the association rule, namely support, confidence, and correlation. The support and confidence parameters are used only for data associations that generate rules for determining methods.

The iteration process is carried out until there are no more combinations or patterns that can be formed, where the results of iteration (1) will be used to calculate the support value for an item.

$$Support\ (\mathrm{A}) = \frac{\text{Number of Transaction A}}{\text{Total Transactions}}\ x\ 100\% \quad (1)$$

Iteration (2) is used for the count score *support* of 2 items.

$$Support\ (\mathrm{A}, \mathrm{B}) = \frac{\text{Number of Transaction A and B}}{\text{Total Transactions}}\ x\ 100\% \quad (2)$$

If there are no more patterns that can be formed, then the results of the iteration will be used to calculate the confidence value of the rule A → B

$$Confidence\ (\mathrm{A} \rightarrow \mathrm{B}) = \frac{\text{Number of Transaction A and B}}{\text{Number of Transaction A}}\ x\ 100\ \% \quad (3)$$

Confidence is a certainty value, namely the strength of the relationship between items in an apriori. Confidence can be searched after the frequency pattern of an item appears is found.

D. Content-Based Filtering

In the implementation of this final project using a content-based filtering approach to support the a priori algorithm method in finding the pattern of relationships between itemsets. Content based filtering provides recommendations based on a similarity in the attributes of the item or items. The system will make an assessment based on the analysis of the similarity of the user profile with the vector components that make up the item. If the item is liked by the user, it will be recommended to the user. The disadvantage of this method is that this method is not able to recommend new types of items or items that have never appeared to users. This is because the content-based filtering method is based on items that have been rated by users or items that are frequently searched by users [4].

III. RESEARCH METHOD

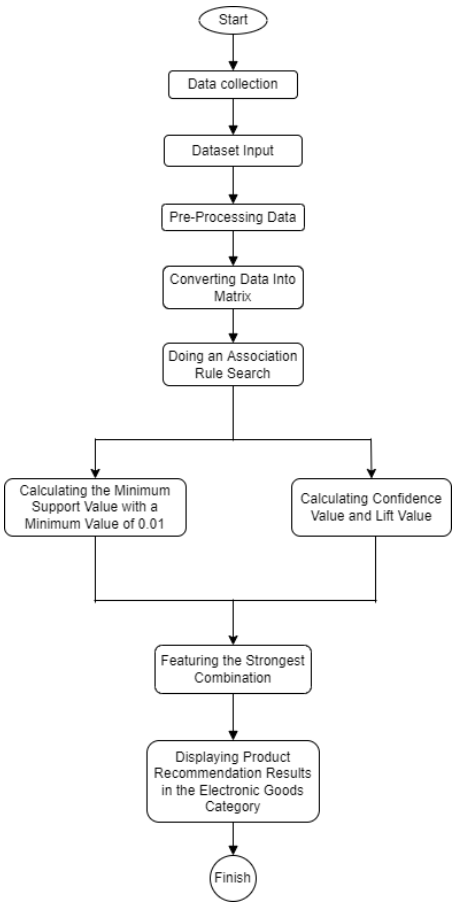

Fig 1. Flowchart Recommender System

The implementation and design of the recommender system that will be used, namely, the first is to input the data that has been searched, after entering the data, it is continued by preprocessing the data to see if there is a missing value and uniting all the datasets used, after the combination has been carried out on the dataset. , then the next step is to change the shape of the data into a vector matrix to make it easier to implement the apriori algorithm on the dataset. After converting the dataset into a matrix form, an apriori algorithm will be implemented by finding the minimum support value with a minimum support value of 0.01 or 1%, after finding the value, it will display the strongest combination, then it will display the results of the recommendation for the electronic goods category data product.

A. Design Implementation of Association Rules

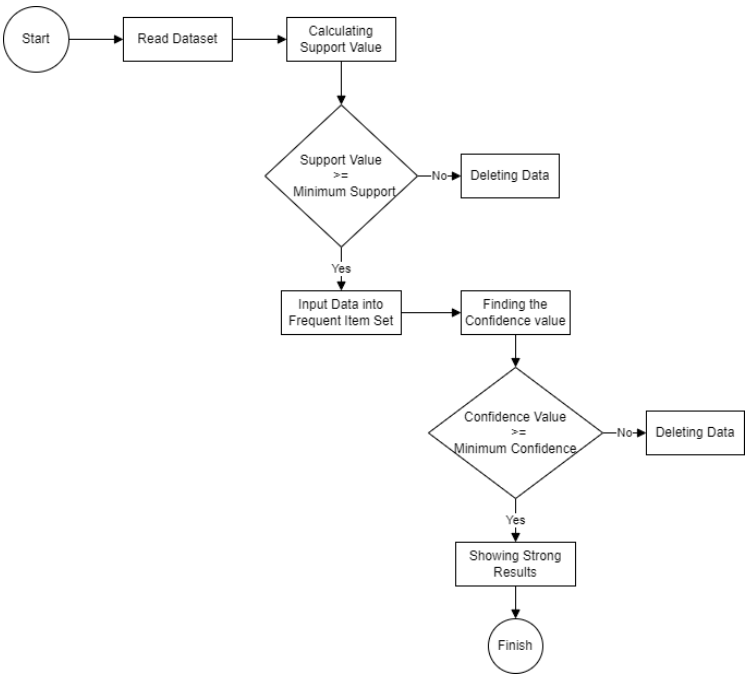

Fig 2. Design Implementation of Association Rules

The design stage of the implementation of association rules, which starts from reading the dataset, then calculating from the support value, if the support value is less than the minimum support value, the data will be deleted, while the support value is large and equal to the minimum support value, the data will be inputted into the frequent itemset, Then the data in the frequent itemset will be searched for the Confidence value, if the confidence value is less than the minimum confidence value, the data will be automatically deleted, while the confidence value is large and equal to the minimum confidence value, the data will be displayed, and display strong results.

B. Apriori Algorithm Implementation Design

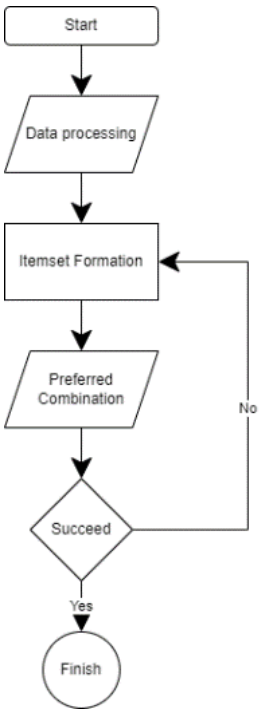

Fig 3. Design Inplementation of Algorithm Apriori

In Figure 3 above, it can be seen that the stages of implementing the a priori algorithm are processing data, then forming the itemset, and finally taking the selected combination, if it is not successful, it will repeat the itemset formation.

IV. Results And Discussion

A. Data Collection

The dataset retrieval of datasets in the development of an e-commerce recommender system on this shopee platform from Kaggle Dataset, where the Kaggle dataset is a web that contains a collection of datasets used in the implementation of a study. The dataset taken is data related to data on the shopee platform in the form of electronic purchase transaction data (Purchase) and electronic search transaction data (Events). The first dataset is the purchase dataset in table I. which contains data for event_time, order_id, product_id, category_id, category_code, brand, price, and user_id with a total data of 2,633,520. The second dataset is the events dataset in table II.which contains data for event_time, event_type, product_id, category_id, category_code, brand, price, user_id, and user_session with a total data of 885,128. The two datasets will be combined in the data pre-processing process and will serve as a process in the e-commerce recommender system.

TABLE I. Table of Purchase Dataset

	event_time	order_id	product_id	category_id	category_code	brand	price	user_id
2633518	2020-11-21 10:10:30 UTC	2388440981134693944	1515966223509089917	2.268105e+18	appliances.personal .scales	vitek	12.48	1.515916e+18
2633519	2020-11-21 10:10:30 UTC	2388440981134693944	2273948184839454837	2.268105e+18	NaN	Moulinex	41.64	1.515916e+18
2633520	2020-11-21 10:10:30 UTC	2388440981134693944	1515966223509127566	2.268105e+18	appliances.kitchen .blender	Redmond	53.22	1.515916e+18

TABLE II. Table of Events Dataset

	event_time	event_type	product_id	category_id	category_code	brand	price	user_id	user_session
885126	2021-02-28 23:58:09 UTC	view	4170534	2144415939364389423	electronics.clocks	amazfit	64.92	1515915625611024020	xNIJBqZdkd
885127	2021-02-28 23:58:14 UTC	view	888273	2144415921932861531	electronics.telephone	NaN	10.16	1515915625611024030	9pCbKMicSx
885128	2021-02-28 23:59:09 UTC	view	743182	2144415935631458761	construction.tools.soldering	Kada	65.08	1515915625556087775	BejOXRngEW

B. *Preprocessing*

At the preprocessing stage, separate data and take some data and then perform missing values, after performing missing values, combine the two datasets into one by taking data categories in the form of category_code and user_id, thus obtaining a total of 937,067 data, and the following is the result of combining the two data the:

	category_code	user_id
0	electronics.tablet.samsung	1.515916e+18
1	electronics.tablet.samsung	1.515916e+18
2	electronics.audio.headphone.huawei	1.515916e+18
3	electronics.audio.headphone.huawei	1.515916e+18
4	furniture.kitchen.table.maestro	1.515916e+18
...	...	...
937062	computers.components.cdrw.asus	1.515916e+18
937063	computers.components.videocards.msi	1.515916e+18
937064	electronics.video.tv.starwind	1.515916e+18
937065	electronics.clocks.amazfit	1.515916e+18
937066	construction.tools.soldering.kada	1.515916e+18

937067 rows × 2 columns

Fig 4. Combine The two Datasets

The next stage is to rearrange the User ID, so that the data on the user id becomes sequential by using the encoder label, along with the data after rearranging the User ID .

	category_code	user_id
0	electronics.tablet.samsung	3828
1	electronics.tablet.samsung	3828
2	electronics.audio.headphone.huawei	11957
3	electronics.audio.headphone.huawei	11957
4	furniture.kitchen.table.maestro	15207
...	...	...
937062	computers.components.cdrw.asus	186442
937063	computers.components.videocards.msi	234223
937064	electronics.video.tv.starwind	234224
937065	electronics.clocks.amazfit	234224
937066	construction.tools.soldering.kada	163846

937067 rows × 2 columns

Fig 5. Rearrange the User ID

The next stage is to group Category based on User, and retrieve data that has transactions more than 1 time, so as to get 141,248 data. And here are the data obtained.

user_id	category_code
1	[electronics.audio.acoustic.logitech, electron...
3	[electronics.telephone.honor, electronics.tele...
5	[computers.peripherals.wifi.zyxel, computers.p...
6	[electronics.telephone.alcatel, electronics.te...
8	[stationery.cartrige.cactus, stationery.cartri...
...	...
234213	[electronics.tablet.huawei, stationery.cartrig...
234216	[computers.components.videocards.msi, computer...
234217	[electronics.video.tv.jvc, electronics.video.t...
234220	[computers.components.motherboard.gigabyte, co...
234224	[electronics.video.tv.starwind, electronics.cl...

141248 rows × 1 columns

Fig 6. Grouping Category Based on User

Then the last stage is to change the shape of the data which was originally in the form of columns into a matrix and vector form using a vectorizer, and the following is the result of the data that has become a vector matrix.

	accessories_bag_airline	accessories_bag_apacer	accessories_bag_arena	accessories_bag_arktika
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
...	...	...	...	...
141243	0	0	0	0
141244	0	0	0	0
141245	0	0	0	0
141246	0	0	0	0
141247	0	0	0	0

141248 rows × 2945 columns

Fig 7. Results of the Data That Has Become a Vector Matrix

C. Implementation of Apriori Algorithm

Based on the results of preprocessing the data, a vector matrix is obtained, where the vector-matrix will determine the frequent itemset by applying the minimum support value = 0.01. A frequent itemset is an item set that has a support value that is greater than or equal to the minimum support value. The frequent itemset in table III. is the result of itemset which has a support value greater than or equal to 0.01.

TABLE III. Frequents Itemsets

	support	*itemsets*
0	0.013359	(appliances_environment_air_conditioner_ava)
1	0.012928	(appliances_environment_air_conditioner_beko)
2	0.023073	(appliances_environment_vacuum_samsung)
3	0.011639	(appliances_environment_vacuum_tefal)
4	0.010740	(appliances_kitchen_kettle_ava)
...	...	...
73	0.012666	(electronics_smartphone_samsung, electronics_s...
74	0.010174	(electronics_smartphone_oppo, electronics_smar...
75	0.012793	(electronics_smartphone_xiaomi, electronics_sm...
76	0.010344	(electronics_video_tv_samsung, electronics_sma...
77	0.010273	(stationery_cartrige_nv, print)

The result of the apriori algorithm search is the calculation of association rules, namely the rules that connect antecedents and consequents. Antecedents are items of the rule, while consequents are items that are related to the rules. Antecedents and Consequents are items that are interrelated and related to each other. An example is "Antecedents = Smartphone, Consequents = Smartphone Case, if someone buys a smartphone then that person will buy a smartphone case to meet their needs, on the other hand if someone buys a smartphone case then that person must buy their smartphone first". This study uses the parameter association rules with a minimum value of confidence = 0.01 and a minimum value of lift = 1. The results of implementing the association rule can be seen in table IV.

TABLE IV. Results of Association Rules

	antecedents	*consequents*	*Support*	*confidence*	*lift*
30	(stationery_cartrige_nv)	(print)	0.010273	1.000000	97.345279
31	(print)	(stationery_cartrige_nv)	0.010273	1.000000	97.345279

From the results of the implementation of association rules on each itemset with the minimum value of support = 0.01, the minimum value of confidence = 0.01, and the minimum

value of lift = 1, it is found that the itemset that produces association rules with the highest accuracy value is itemset stationery_cartrige_nv and print where the itemset has a support = 0.01, confidence value = 1.00, and lift value = 97.34. The itemset can be said to have the highest accuracy value because it has a very high confidence value and lift value and the itemset between antecedents and consequents are related to each other and have a relationship pattern or are called mutual combinations, resulting in a high confidence value. These results can be interpreted as follows "If someone buys or looks for a stationery_cartrige_nv then that person will also buy or will look for prints on the activities they are doing, and vice versa". So it can be concluded that the higher the support value and the confidence value from the minimum value, the higher the accuracy value obtained. The results of the calculation evaluation show that the use of apriori algorithms is suitable in predicting a recommendation and can improve performance on the results obtained.

V. Conclusion

Based on the results of testing and analysis on the development of an e-commerce recommender system using the apriori algorithm, it provides better recommendation results than without using the a priori algorithm or only using content-based filtering. The apriori algorithm also provides the association rule value that has the highest accuracy and increases the user's confidence value for the recommended item.

The results of this study indicate that the association rules with the highest accuracy values are itemset stationery_cartrige_nv and print where the itemset has a support value = 0.01, a confidence value = 1.00, and a lift value = 97.34. The itemset can be said to have the highest accuracy value because, it has a very high confidence value and lift value and the itemset between antecedents and consequents are related to each other and have a combination of each other, resulting in a high confidence value and the application of the apriori algorithm can increase confidence value for the recommended item.

In future research, improve the performance of the recommender system with a larger e-commerce dataset using different parameters. Then combine the apriori algorithm with the addition of other methods to create a system of recommenders.

Acknowledgment

The author would like to thank especially my supervisor Mr. Z K Abdurahman Baizal and also friends who have supported the author in designing this recommendation system.

References

[1] Ching, K. C., Hasan, Z. R., & Hasan, N. A. (2021). Factors influencing consumers in using shopee for online purchase intention in east coast Malaysia. UMT Journal of Undergraduate Research, 3(1), 45-56.
[2] Baizal, Z. A., Widyantoro, D. H., & Maulidevi, N. U. (2016, October). Query refinement in recommender system based on product functional requirements. In 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 309-314). IEEE.
[3] Baizal, Z. K. A., Widyantoro, D. H., & Maulidevi, N. U. (2020). Computational model for generating interactions in conversational recommender system based on product functional requirements. Data & Knowledge Engineering, 128, 101813.
[4] Chandra, A., Ahmed, A., Kumar, S., Chand, P., Borah, M. D., & Hussain, Z. (2022). Content-Based Recommender System for Similar Products in E-Commerce. In Edge Analytics (pp. 617-628). Springer, Singapore.
[5] Koesmyta, R. P. (2020). Pembuatan E-Commerce Berbasis Web Dengan Sistem Rekomendasi Content-Based Filtering Menggunakan Algoritma Apriori (Doctoral dissertation, Universitas 17 Agustus 1945 Surabaya).

[6] Badriyah, T., Fernando, R., & Syarif, I. (2018). Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. Konferensi Nasional Sistem Informasi (KNSI) 2018.
[7] Riyadi, N., Mulki, M. F., & Susanto, R. (2019). Analysis of Customers Purchase Patterns of E-Commerce Transactions Using Apriori Algorithm and Sales Forecasting Analysis With Weighted Moving Average (WMA) Method. Scientific Research Journal.
[8] WANG, X. Y. (2017). An Apriori-Based E-Commerce Recommendation System. DEStech Transactions on Economics, Business and Management, (icem).
[9] Pradana, H. A., Laurentinus, F. P. J., & Sylfania, D. Y. (2020). Product Recommendation Systems using Apriori in the Selection of Shoe based on Android.
[10] Vishwakarma, G., Parte, S., Joshi, H., Patel, J., & Jangid, P. (2019). Online Sequential Behaviour Analysis using Apriori Algorithm.
[11] Fatoni, C. S., Utami, E., & Wibowo, F. W. (2018, December). Online Store Product Recommendation System Uses Apriori Method. In Journal of Physics: Conference Series (Vol. 1140, No. 1, p. 012034). IOP Publishing.
[12] J. Han, M. Kamber and J. Pei, 2012, Data Mining: Concepts and Techniques Third Edition, Elsevier
[13] Kusrini, Luthfi, E. T., 2009, Algoritma Data Mining, Andi Offset, Yogyakarta