Product Recommendation Systems using Apriori in the Selection of Shoe based on Android

Harrizki Arie Pradana[1], Laurentinus[1], Fransiskus Panca Juniawan[1] and Dwi Yuny Sylfania[1]
[1]*Department of Informatics Engineering, STMIK Atma Luhur, Jl. Jend. Sudirman-Selindung, Pangkalpinang-Bangka,Indonesia*

Keywords: Recommendation Systems, Apriori, Rules, Consequent, Android.

Abstract: The development of information technology has given a lot of influence and changes in human life. One of the areas of life affected is the field of trade. Currently, there are still many people who are looking for shoes manually by going around looking for shoe stores, and the results are wasting time, energy, and transportation costs. In addition, by buying manually the buyer does not know the experience of other buyers who have bought shoes with the same model. In addition, by implementing the recommendation system using the Apriori algorithm, prospective buyers can find out the experience of previous buyers. The advantage of the Apriori algorithm is simpler, more efficient, and able to handle large amounts of data than other algorithms. This study uses the waterfall model with four stages in it. The system was tested using the Apriori method to produce rules derived from the pattern of the combination of two items, rules and consequent. The results of this study are the successful search for the level of support and trust in shoes so that the shoes obtained are more accurate using data mining.

1 INTRODUCTION

The development of information technology in the field of communication has at least two rapidly developing technologies. The first is a cellular telephone and the second is an internet networked computer (Kasemin et al., 2016). Mobile phones are no longer only used for a phone call and short messages but have developed into multifunctional devices or what we know better as smartphones.

Business processes in the field of trade cannot separate from the use of smartphones based on Android. Android is developed based on the Linux kernel system so that it categorizes in an open operating system. Most of the leading cellphone vendors are currently using Android as an operating system (OS) for their smartphones. Android has transformed into the most widely used operating system for smartphones in just a few years after its appearance (Kurniawati et al.,). However, user privacy remains a key aspect of every mobile application (Aceto et al., 2018)(Aceto et al., 2019).

Shoes, which are one of the human needs, have increased in use over time, and have become an obligation in various fields of work, education and a fashion trend which has increased the number of models and brands of shoes from year to year. The process of finding shoes for some people is still doing manually. Manually it means that when you want to buy new shoes, they need to get around in search of a shoe store that is about to sell shoes they like. This process will cause waste of time and transportation costs which can reduce if the community is not confused in finding a shoe store (Badriyah et al., 2018).

One solution to overcome the above problems is to make a mobile application based on android shoes. Android choose because it has many advantages over platforms (desktop, website) and operating systems (OS) smartphones (such as iOS, Windows Phone, Symbian). The benefits of Android include large market share, open source, easy to use compared to other platforms and OS, and easy to use anytime and anywhere.

A mobile application that is equipped with recommended algorithm will help the user to determine their choice. There are many algorithms that can be used to make recommendations, in this case the purchase of shoes. One of them is Apriori. This study uses Apriori because of its advantages in doing generates candidate items set and tests if they are frequent (Kavitha and Selvi, 2016).

Applications that made later can be used by cus-

Pradana, H., Laurentinus, ., Juniawan, F. and Sylfania, D.
Product Recommendation Systems using Apriori in the Selection of Shoe based on Android.
DOI: 10.5220/0009090603110318
In *Proceedings of the International Conferences on Information System and Technology (CONRIST 2019)*, pages 311-318
ISBN: 978-989-758-453-4

tomers to check product recommendations, prices, and shoe details in the shop which can then be a reference for customers in choosing the desired shoes. Product recommendation system with Apriori algorithm will add to the application. The product recommendation system allows applications to display products that might be liked by users, thereby reducing product search time. Thus, users can make their choices according to the trend of shoe users. In addition, the choice can be accelerated so they can save their resources.

2 LITERATURE REVIEW

2.1 Recommendation Systems

The recommendation system is an application that provides and recommends an item in making a decision desired by the user. The implementation of recommendations in order usually predicts an object, such as film and music recommendations. The system runs in two ways, namely by collecting user data directly and indirectly.

Direct data collection is asking the user to rate an item. While indirect data collection is by observing the objects, then it could be seen by users on an ecommerce web. After the observational data collected, then it is processed using a particular algorithm. After that, the results will return to 10 users as an item recommendation with the parameters of that user. The recommendation system is also an alternative in searching for an item sought by users (Raharjana, 2017).

2.2 Apriori Algorithm

Apriori is part of the association rule method, which serves to find item combinations based on items purchased by customers. Types of association rules include a priori, generalized rule induction, and hash-based algorithms. Association rule mining is a data mining technique to find associative rules between a combination of items, for example, analysis of purchases at a supermarket. With the existence of data and observations, it can be known some possibilities for a customer to buy bread together with milk.

By utilizing this condition, self-service owners can take advantage of these conditions by regulating the placement of goods or designing marketing promotions (Febriansyah and Samsinar, 2018).

There are three stages to determine frequent patterns (Kavitha and Selvi, 2016), namely:

2.2.1 Generate and Test

The first step is to determine the 1-itemset frequent L1 elements by scanning the database. Then remove all elements from C that do not meet the minimum criteria.

2.2.2 Join Step

To reach the element at the next level, Ck joins the frequencies of the previous elements by self join.

Suppose that Lk-1 * Lk-1 is known as a Cartesian product from Lk-1. This stage generates new candidate k-itemsets based on combining Lk-1 with them which was found found in the previous iteration. Then Ck denote candidate k-itemsets and Lk becomes the frequent k-itemset.

2.2.3 Prune Test

Prunning eliminates several candidates from the kitemset using the Apriori principle. The database scanning process is carried out to determine the number of each candidate in Ck which will result in the determination of Lk (that all candidates have an amount less than the minimum amount of support). Repeat steps 2 and 3 until no more sets of new candidates are generated.

2.3 Previous Study

There have been many previous studies using Apriori algorithms. There is study uses the itembased collaborative filtering method, where the system will look for similarities in purchasing models with others. Next, the system will search for ratings between items based on the level of similarity that exists. After the evaluation between pieces is obtaining, then this rating will be calculated similarity value between objects using the Adjusted Cosine Similarity approach. The results of the similarity calculations between items will use for the next stage. This stage predicts a rating that has never been done by a customer for a particular subject. This approach uses the Weighted Sum formula whose prediction value will make a recommendation to the customer (Kurniawan, 2016). The application of a priori algorithm for movie website recommendations is done by using a new approach to adjust the features displayed and have an impact on increasing the representative of the movie (Ma, 2016) (Pal et al., 2017).

Apriori and Content-Based Filter (CBF) is also used for determining the supply of compressor spare parts (Kurniawati et al.,), market basket analysis in the mini market (Elisa et al., 2018)(Mauliani et al.,

2015), sales analysis (Nursikuwagus and Hartono, 2016), positioning of products (Ningsih et al., 2016), property search (Badriyah et al., 2018), game's hardware (Yanti et al., 2015), and product recommendation (Putra et al., 2019)(Kurniawan, 2016).

Also conducted research to design step-bystep models to study the predictive analysis of Apriori algorithms. From the weakness, namely speed, several methods are proposed to overcome it, namely dynamic counting, determining the largest value for support and confidence, and also taking large amounts of dataset sample data (Roşca and Rădoiu, 2015).

Apriori can also be combined with other methods. Combined Apriori with binary k-means at the key parameter operating distance in the Central Air Conditioning System (CACS). The results of this study prove that priori discovered the rules of energy consumption. Operating data from equipment was analyzed by DM, and cluster analysis was used to improve the control effects of CACS (Yan et al., 2019). Furthermore, there are Apriori combinations and clustering algorithms that are used to get datasets from traffic accidents. With the help of a tool from WEKA which provides many algorithms for data mining selected a priori and clusters. From the test results, it is proved that Apriori is better than the EM cluster (Nafie Ali and Mohamed Hamed, 2018).

Content-Based Filter (CBF) and Content Filtering (CF) combination were also conducted as hybrid mechanism filters that can remove recommendations that are not relevant to search keywords made by mobile-based users. From the test results it is proven that the proposed filtering mechanism can improve user personalization and enhance the filtering experience on mobile devices (Zhao et al., 2015). Apriori and Content-Based Filter (CBF) are not only used for recommendations. There is research on improving network sensors by modifying the priori algorithm. From the test results it is known that modified algorithms can better combine Wireless Sensor Network (WSN) models from mobile nodes, also get greater network coverage (Ji and Zhang, 2018). Then a study was conducted to improves GPU performance with the Apriori algorithm. From the test results it was proven that GPU Apriori was able to improve the efficiency of item set mining (Jiang et al., 2017).

3 RESEARCH METHODOLOGY

This research consists of four stages, namely needs analysis, algorithm analysis, implementation, and testing.

3.1 Needs Analysis

At this stage, we do some data collection and needs analysis related to the system being built is carried out. The data is collected from books and articles from journals or proceedings. Needs analysis includes the specifications of the system to be built and the computer used.

3.2 Algorithm Analysis

At this stage we learn and find the ways of working, how to calculations, and learn the flow process of the Apriori algorithm that implemented in the mobile apllication.

3.3 Implementation

The results of the learning of the Apriori algorithms are then applied in the Android system. To help to understanding, case studies were given on the android-base shoe ad application.

3.4 Testing

After the algorithm is applied, then the application is tested to determine the performance of the Apriori algorithms that implemented and performance of the application. The tests carried out are algorithm performance testing using whitebox and blackbox method.

4 RESULTS AND DISCUSSIONS

4.1 Problems Analysis

The types and brands of shoes in shoe stores continue to increase each year to meet the different interests of customers. But with the varying number of shoe products, customers need a long time to choose the desired shoes.

4.2 Proposed System Analysis

The explanation of the proposed system analysis will be explained using use case diagrams, activity diagrams, and class diagrams, which can see below:

4.2.1 Use Case Diagram

Based on Figure 1, it can explain that in the system there are two actors, namely the user and admin.

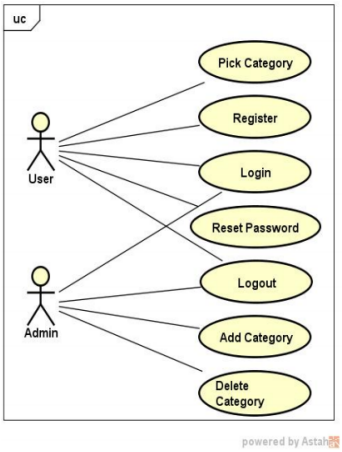

Figure 1: Use case diagram

4.2.2 Activity Diagram

An activity diagram is a diagram that describes the flow of functionality from the system. The types of activity diagrams this time are registers, logins, change passwords, log out, add advertisements, delete ads, add categories, delete categories, and activity diagrams select categories. For example, can be seen in Figure 2, for activity logins diagrams.

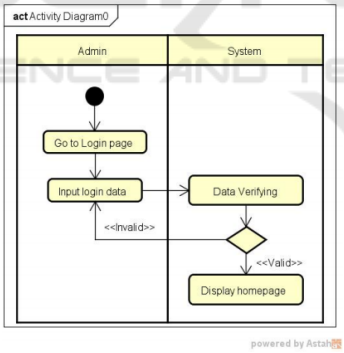

Figure 2: Login processes Activity Diagram

4.2.3 Class Diagram

Class diagrams are diagrams to display several classes in the software system that will develop. The proposed class diagram can see in Figure 3.

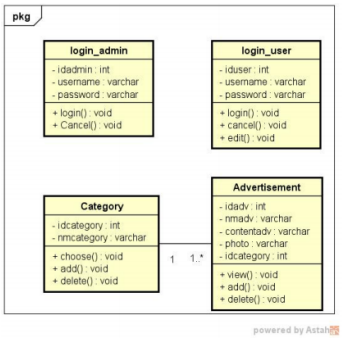

Figure 3: Class diagram

4.3 Discussion of Apriori Algorithm

Apriori algorithms are using so that computers can learn the rules of the association, look for patterns of relationships between items in data so that they can make a product recommendation. An example of transaction data can see in Figure 4.

Product	2018											
	1	2	3	4	5	6	7	8	9	10	11	12
Nike		1	1		1			1		1		1
Adidas	1	1	1	1	1		1	1	1		1	1
New Balance	1		1	1	1				1	1	1	1
Dior		1		1			1			1		1
Puma					1	1		1				

Figure 4: Transaction data.

The steps of using a priori algorithm can see in the explanation below:

4.3.1 First Iteration

In first iteration, calculate the number of transactions for each item. Where the method of calculating support can see in equation (1).

$$Support(A) = \frac{Numberoftransactionscontaining(A)}{TotalTransactions} x100 \quad (1)$$

The results of calculations in first iteration can be seen in Figure 5.

Item set	Support count (Total 12)	Support
Adidas	10	83.3%
New Balance	8	66.7%
Nike	6	50%
Puma	3	25%
Dior	5	41.7%

Figure 5: First iteration results.

4.3.2 Second Iteration

In second iteration, do a combination of the previous k-itemset, where the method of calculating support can see in equation (2).

$$Support(A,B) = P(A \cap B) = \frac{Numberoftransactioncontaining(AandB)}{TotalTransactions} x100 \quad (2)$$

The results of calculations in the second iteration can see in Figure 6. Support Count is obtained from shoe transaction data that have been collected safely in 2018, namely 24 transactions. Then it is sought a transaction involving two brands of shoes. In the elaboration of Figure 6, the first point mentioned that Adidas and Nike have a support count of 8 transactions. This means that out of the 24. transactions there are 8 transactions that intersect between Adidas and Nike.

Item set	Support	Support count (Total 24)
Adidas → Nike	33.3%	8
Adidas → New Balance	58.3%	14
Nike → New Balance	33.3%	8
Nike → Puma	16.7%	4
Adidas → Dior	33.3%	8
New Balance → Dior	8.3%	2
Adidas → Puma	16.7%	4
Nike → Dior	25%	6

Figure 6: Second iteration results.

4.3.3 Establishment of Associative Rules

Calculate confidence using a formula that can see in equation 3.

$$Confidence = P(B|A) = \frac{Numberoftransactioncontaining(AandB)}{Totaloftransactioncontaining(A)} x100 \quad (3)$$

The results are shows in Figure 7.

Item Set	Confidence
Adidas -> Nike	45.45%
Nike -> Adidas	83.3%
Adidas -> New Balance	63.6%
New Balance -> Adidas	87.5%
Nike -> New Balance	66.7%
New Balance -> Nike	50%
Nike -> Puma	33.3%
Puma -> Nike	66.6%
Adidas -> Dior	36.3%
Dior -> Adidas	60%
New Balance -> Dior	37.5%
Dior -> New Balance	60%
Adidas -> Puma	18.2%
Puma -> Adidas	66.6%

Figure 7: Associative rules results.

4.4 Implementation

Screen design is the first display that is in the application for the admin. The following Figure 8 is a screen design minimum support for database sales.

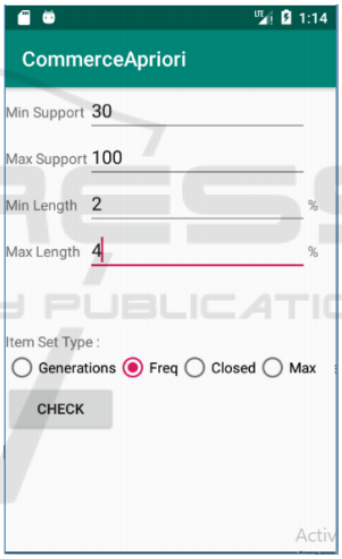

Figure 8: Minimum Support for database sales

Another screen design, can see in Figure 9, it is about item formation results from the database. Figure 10 shows a display of determining minimum parameters and Figure 11 is about the results of the association carried out from Apriori algorithm.

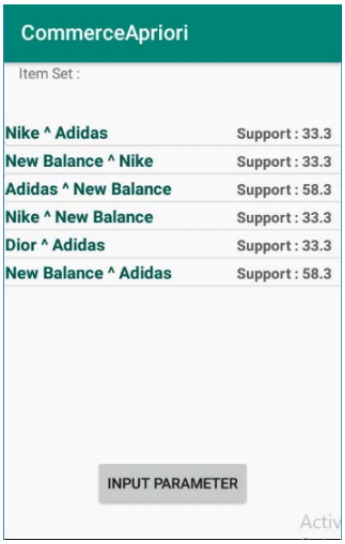

Figure 9: Item formation results from the database

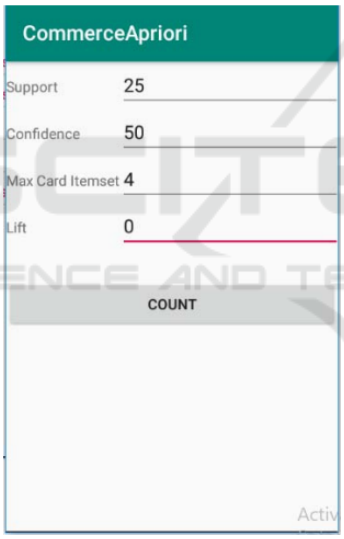

Figure 10: display of determining minimum parameters

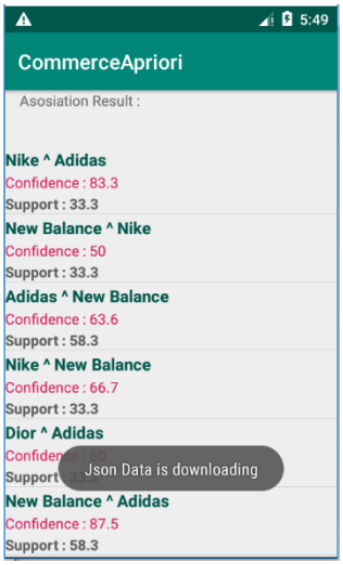

Figure 11: display of the results of the association carried out a priori based.

4.5 Testing

To find out the performance of the application and the algorithm that has been made, then testing is done using Whitebox and Blackbox method. The whitebox used to testing based on manual calculation and the application calculation as shows in Figure 12, and the Blackbox testing used to measure the perform of the application and shows in Figure 13.

Both testing methods are considered to have represented the purpose and utilization of the recommendations made.

Testing Item	Manual Result	App Results	Testing Result
Adidas -> Nike	45.45%	45.45%	Valid
Nike -> Adidas	83.3%	83.3%	Valid
Adidas -> New Balance	63.6%	63.6%	Valid
New Balance -> Adidas	87.5%	87.5%	Valid
Nike -> New Balance	66.7%	66.7%	Valid
New Balance -> Nike	50%	50%	Valid
Nike -> Puma	33.3%	33.3%	Valid
Puma -> Nike	66.6%	66.6%	Valid
Adidas -> Dior	36.3%	36.3%	Valid
Dior -> Adidas	60%	60%	Valid
New Balance -> Dior	37.5%	37.5%	Valid
Dior -> New Balance	60%	60%	Valid
Adidas -> Puma	18.2%	18.2%	Valid
Puma -> Adidas	66.6%	66.6%	Valid

Figure 12: Whitebox Testing Manual Calculation.

Testing Item	Results	Validation
Radio Button in item set type	Data saved based on chosed radio button	Valid
Check Button in item set type	Display support page	Valid
Input Parameter Button in Item Set	Display input parameter page	Valid
Count Button	Display recommendation page	Valid

Figure 13: Blackbox Testing.

5 CONCLUSIONS

The results of the study are looking for the level of support and confidence in shoes so that shoes are obtained that are more accurate using data mining. Based on testing the system uses the apriori method to produce rules derived from the pattern of the combination of two items. From the calculation results obtained that the highest associative rule product is New Balance → Adidas with a value of 87.5% and the lowest result on the Adidas → Puma with a value of 18.2%. The rules above consist of Antecedent is a form of a condition rather than rules, a consequent is a form of a statement rather than rules, lift shows the level of power rules random events of antecedent and consequent based on their respective support.

Further research can carry out a variety of additional testing to test the performance of the apriori algorithm that created and implemented. Besides that, it can also combine or compare apriori algorithms with other recommendation algorithms.

REFERENCES

Aceto, G., Ciuonzo, D., Montieri, A., and Pescapé, A. (2018). Multi-classification approaches for classifying mobile app traffic. *Journal of Network and Computer Applications*, 103:131–145.

Aceto, G., Ciuonzo, D., Montieri, A., and Pescapé, A. (2019). Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges. *IEEE Transactions on Network and Service Management*, 16(2):445–458.

Badriyah, T., Azvy, S., Yuwono, W., and Syarif, I. (2018). Recommendation system for property search using content based filtering method. In *2018 International Conference on Information and Communications Technology (ICOIACT)*, pages 25–29. IEEE.

Elisa, E. et al. (2018). Market basket analysis pada mini market ayu dengan algoritma apriori. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(2):472–478.

Febriansyah, R. and Samsinar, S. (2018). Desain sistem informasi pembelian dan penjualan dengan unified modeling language (uml) tools pada toko simplebag. *IDEALIS: InDonEsiA journaL Information System*, 1(1):7–11.

Ji, Q. and Zhang, S. (2018). Research on sensor network optimization based on improved apriori algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2018(1):1–9.

Jiang, H., Xu, C.-W., Liu, Z.-Y., and Yu, L.-Y. (2017). Gpu-accelerated apriori algorithm. In *ITM Web of Conferences*, volume 12, page 03046. EDP Sciences.

Kasemin, H. K. et al. (2016). *Agresi Perkembangan Teknologi Informasi*. Prenada Media.

Kavitha, M. and Selvi, S. T. (2016). Comparative study on apriori algorithm and fp growth algorithm with pros and cons. *International Journal of Computer Science Trends and Technology (I JCS T)–Volume*, 4.

Kurniawan, A. (2016). Sistem rekomendasi produk sepatu dengan menggunakan metode collaborative filtering. In *Seminar Nasional Teknologi Informasi dan Komunikasi*, pages 610–614.

Kurniawati, L., Kusuma, A. E., and Dewansyah, B. Implementasi algoritma apriori untuk menentukan persediaan spare part compressor. *Computer Engineering, Science and System Journal*, 4(1):6–10.

Ma, K. (2016). Content-based recommender system for movie website.

Mauliani, A., Hartati, S., and Musdholifah, A. (2015). Pembentukan temporal association rules menggunakan algoritma apriori (studi kasus: Toko batik diyan solo). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 10(1):71–80.

Nafie Ali, F. M. and Mohamed Hamed, A. A. (2018). Usage apriori and clustering algorithms in weka tools to mining dataset of traffic accidents. *Journal of Information and Telecommunication*, 2(3):231–245.

Ningsih, W. et al. (2016). Implementasi data mining untuk penentuan posisi barang pada rak menggunakan metode apriori pada pt midi utama indonesia. *Techno. Com*, 15(4):335–342.

Nursikuwagus, A. and Hartono, T. (2016). Implementasi algoritma apriori untuk analisis penjualan dengan berbasis web. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 7(2):701–706.

Pal, A., Parhi, P., and Aggarwal, M. (2017). An improved content based collaborative filtering algorithm for movie recommendations. In *2017 tenth international conference on contemporary computing (IC3)*, pages 1–3. IEEE.

Putra, A. A. C., Haryanto, H., and Dolphina, E. (2019). Implementasi metode association rule mining dengan algoritma apriori untuk rekomendasi promo barang. *CSRID (Computer Science Research and Its Development Journal)*, 10(2):93–103.

Raharjana, I. K. (2017). *Pengembangan Sistem Informasi Menggunakan Metodologi Agile*. Deepublish.

Rosca, D. G. and Radoiu, D. (2015). Step-by-step model for the study of the apriori algorithm for predictive analysis. *Scientific Bulletin of the Petru Maior" University of Targu Mures*, 12:44–47.

Yan, L., Qian, F., and Li, W. (2019). Research on key parameters operation range of central air conditioning based on binary k-means and apriori algorithm. *Energies*, 12(1):102.

Yanti, N., Rahmi, R., and Ruliah, R. (2015). Penerapan algoritma collaborative filtering untuk rekomendasi games hardware. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 2(1).

Zhao, F., Yan, F., Jin, H., Yang, L. T., and Yu, C. (2015). Personalized mobile searching approach based on combining content-based filtering and collaborative filtering. *IEEE Systems Journal*, 11(1):324–332.