Article Recommendation Using Apriori Algorithm on Website

Ivan Yusuf Rahardika[1], Amak Yunus Eko Prasetyo[2], Moh. Ahsan[3]
Department of Information Technology[1,2,3],
Universitas Kanjuruhan Malang, Indonesia[1,2,3]
ivanyusuuf@gmail.com[1], amakyunusep@unikama.ac.id[2], ahsanm@unikama.ac.id[3]

ABSTRAK

Website adalah kumpulan halaman-halaman situs yang terangkum dalam sebuah domain atau subdomain di internet. Pada umumnya website menampilkan banyak macam artikel yang menarik dan bermanfaat bagi pengunjung. Art7wing merupakan salah satu website yang fokus pada informasi teknologi dan hiburan. Namun, Art7wing hanya menampilkan artikel yang baru dipublish yang membuat artikel lainnya memiliki sedikit kesempatan untuk dikunjungi. Sehingga, untuk memperlihatkan artikel yang lama, maka Algoritma Apriori digunakan. Algoritma Apriori merupakan jenis aturan asosiasi pada data mining. Penggunaan Algoritma Apriori ditujukan untuk menemukan aturan asosiasi dalam rekomendasi artikel dalam Art7wing. Sehingga, penelitian ini memiliki tujuan untuk mengetahui penggunaan Algoritma Apriori dalam memberikan rekomendasi artikel pada website Art7wing. Menggunakan Multimedia Development Life Cycle oleh Luther sebagai metodologi penelitian ini. Multimedia Development Life Cycle terdiri dari 6 tahapan; konsep, desain, pengumpulan bahan, perakitan, percobaan dan distribusi. Hasil penelitian menunjukkan bahwa Algoritma Apriori dapat menampilkan rekomendasi artikel pada Art7wing. Terbukti dengan berhasilnya rekomendasi artikel yang tampil pada halaman artikel website Art7wing.

Keywords: Website; PHP; Algoritma Apriori; Rekomendasi Artikel

1. INTRODUCTION

Website is a group of sites which recapitulated in a domain or subdomain inside World Wide Web (www) in internet. Generally, a website shows the variety of interesting and advantageous content for visitors. However, the amount of contents could affect the room space in a website. Moreover, most of informational website shows a huge number of articles which arranged based on the publication date (Trimarsiah & Arafat, 2017). One of the website which focused on technology information and entertainment is Art7wing. Art7wing aimed to give information and inspiration through its content to visitors. However, since Art7wing only has two website's administrators in maintaining the website, then the article provided there is slight. Similar to other website, Art7wing showed only recent published article which made the other articles had little chance to be visited or read by visitors.

Therefore, in order to solve the problem, an effort to show the recommended linked article is suggested. This effort could open opportunity for other articles to be read by visitors. Showing recommended article could be done using Apriori Algorithm technique. Apriori Algorithm is a kind of association rule in data mining. The rule which reveals the association between some attributes is often called as affinity analysis or market basket analysis. Association rule mining is used to define the associative rule between item combinations. The importance of association rule could be seen in two parameters; support and confidence. Support is an item combination presented inside a database, while confidence is a strong relationship between items inside association rule (Kusrini & Luthfi, 2013). Apriori Algorithm is intended to find the association rules of article recommendation in Art7wing website by mining the relation between items contained inside the visitors' history. Thus, this study aimed to know the implementation of Apriori Algorithm in giving article recommendation in Art7wing website. This study has its scope and limitation in giving some article recommendation link in Art7wing

using Apriori algorithm, using PHP as programming language, having 6000 reading history in last two years It has be contain with at least 5 literature in order to justify novelty this paper.

II. METHODS

2.1. Research Method Development

This study used research and development model with Luther's development model as the methodology. The development model could create a new product which has never existed before and fix the existing product to be more practical, effective and efficient (Sugiono, 2016). Luther's development model or known as Multimedia Development Life Cycle (MDLC) was derived in six steps, those are; concept, design, material collecting, assembly, testing and distribution. These steps are allowed to be put randomly or in unorganized way in its practice. However, even though the steps are allowed to be placed randomly, but the first step must be making the concept (Sutopo, 2012). According to Binanto, MDLC had some advantageous in developing product, such as; (a) easy to understand and implement, (b) the steps are clear and easy to follow, (c) structured and ordered logically, (d) could be used as small development (Binanto, 2013).

2.2. Research Procedure Development

Multimedia Development Life Cycle (MDLC) is a multimedia application which often used as a research development. The MDLC model could be seen in figure 1 below;
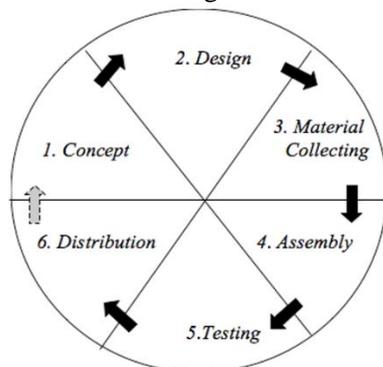

Figure 1. MDLC Methodology (Sutopo, 2012)

1. Concept

This step explained about the website development, including the website's description, purpose and goal. In this study, concept could be used as website development which has article recommendation. This step could be the basic in developing Art7wing website.

2. Design
Design is a step in making specification on website architecture, display and material necessity or material for program, and Apriori Algorithm application for giving articles recommendation. Here, the researcher designed the website operational from the beginning of the website's opening until users or visitors get the recommended articles.

3. Material Collecting
This step is a step in collecting necessity materials for website development. The materials were consisted of picture, text, and other materials which obtained free or self-made from supporting software.

4. Assembly
The assembly step is a step in assembling the materials used for website development until it becomes a website which able to give article recommendation using association rules obtained from Apriori Algorithm.

5. Testing
The next step is testing the software. A prepared software is tested to make sure that all parts has been worked. This step is done to minimalize mistake and ensure the result, whether it is suitable with the expectation or not.

6. Distribution
In this step, the application is saved in a storage which could be reached by users or visitors. The distribution process is consisted of the process of uploading the data to virtual server. The website storage is in the form of folder and file with the variety of access right in order to keep the website safety and developed.

2.3. Developing Early Product

Product development was started from making the concept, designing the product,

collecting the material, and assembling the product. Concept was described as a general description on the development purpose. Then it designed based on the software choices, UML (Unified Model Language), and database and interface design. Then materials for system development were collected. After collecting the materials, then the materials were assembled in a system with programming language thus the function could worked well and used by users or visitors.

2.4. Apriori Algorithm

Algorithm was controlling the development of item set candidate and the result of frequent item set using support-based pruning to omit the unattractive item set by setting the minimum support (Fadlina, 2014). Apriori Algorithm was one of the association rule in data mining. Besides Apriori Algorithm, generalization rule induction and algorithm hash based were also the association rule in data mining. The rule which explained on the association between some attributes was often called as affinity analysis or market basket analysis.

Association analysis or association rule mining is a technique in finding the associative rule between an item combinations. For example in supermarket purchasing analysis, it could be known the possibilities of customer in buying bread together with milk. With this knowledge, the supermarket owner could arrange the placement of his product or design the marketing campaign by giving discount voucher for buying a certain thing. Since this application is popular in analyzing the supermarket basket, then this association analysis was often known as market basket analysis. Association analysis was also known as the basic technique of other data mining. Particularly, one of the step in getting an efficient algorithm in associative analysis was using high frequent pattern mining analysis.

The importance of association rule could be seen in two parameters; support and confidence. Support is an item combination presented inside a database, while confidence is a strong relationship between items inside association rule (Kusrini & Luthfi, 2013). The association rule was often stated as below;

{bread.margarine} ⇨ {milk} (support40%, confidence50%)

The association rule above had meaning on; "50% of transaction in database which contain bread and margarine item along with milk item. While the 40% of whole transaction in database contained those three items."

Association analysis was defined as a process in finding all the association rules which fulfilled the minimum support requirement and minimum confidence requirement. Basic methodology in associative analysis was divided into two phases;

1. High frequent pattern analysis
   This phase was looking for item combination which fulfilled the minimum requirement from support score within the database. The support score of an item was obtained with equality as below;
   $$Support\ (A) = \frac{\sum A\ Transaction}{\sum Transaction}$$
   Meanwhile, the support score of two items were obtained with equality as below;
   $$Support\ (A, B) = P\ (A \cap B)$$
   $$Support\ (A, B) = \frac{\sum A\ and\ B\ Transaction}{\sum Transaction})$$

2. Forming association rule
   After the whole high frequent pattern was found, then the association rule which fulfilled the minimum requirement for confidence by counting the association rule of confidence A ⇨ B could be found. The confidence score from A ⇨ B was obtained with equality below;
   $$Confidence = P\ (B \mid A)$$
   $$Confidence = \frac{\sum A\ and\ B\ Transaction}{\sum A\ Transaction}$$

III. RESULT AND DISCUSSION

3.1. Apriori Algorithm Design

The design of Apriori Algorithm in a developed website was the depiction of Apriori Algorithm application to produce article recommendation using article's history. The first process was taking the visitor history on the website.

Table 1. Visitors' History

Visitor	Category of Article Item

1	News, Music, Game, Computer, Technology, Linux, Entertainment, Social
2	Game, News
3	Game, News
4	Game, News
5	Game, Technology, News, Social
6	Technology, News, Linux
7	Music, Game, Computer, News
8	News, Game, Social, Technology, Music
9	Music, News, Game
10	Linux, Game, News

Table 2. Tabular Format of Visitor History

Visitor	News	Music	Game	Computer	Technology	Linux	Entertainment	Social
1	1	1	1	1	1	1	1	1
2	1	0	1	0	0	0	0	0
3	1	0	1	0	0	0	0	0
4	1	0	1	0	0	0	0	0
5	1	0	1	0	1	0	0	1
6	1	0	0	0	1	1	0	0
7	1	1	1	1	0	0	0	0
8	1	1	1	0	1	0	0	1
9	1	1	1	0	0	0	0	0
10	1	0	1	0	0	1	0	0

The application of Apriori Algorithm was divided into two phases; the analysis of high frequent pattern and forming association rule. Iteration was done toward the development of item combination from item set candidate. In doing the analysis, the minimum support was defined to find the frequent item set.

Table 3. Candidate 1 – item set

Item	Support
News	10/10 100%
Music	4/10 40%
Game	9/10 90%
Computer	2/10 20%
Technology	4/10 40%
Linux	3/10 30%
Entertainment	1/10 10%
Social	3/10 30%

In item set candidate 1, the item combination was not done. The Item was obtained from all article categories showed in visitor's history as seen in table 1 above. Then the support score in each item was counted with equation of (1). For instance, in Computer item, the support score was derived from the total appearance of Computer for every visitor, that was; 2 divided by 10 (the total of visitors) which obtained 0.2 or 20% for the result. If the minimum support was 40% then the frequent item set (F_1) was {News, Music, Game, and Technology}.

Table 4. Candidate 2- item set

Combination of Article Category		Support
A	B	
News	Music	4/10 40%
News	Game	9/10 90%

News	Technology	4/10 40%
Music	Game	4/10 40%
Music	Technology	2/10 20%
Game	Technology	3/10 30%

For item set candidate 2 and the rest item set candidate as seen in table 4, the item of the item set candidate was obtained from frequent item set in previous item set candidate. In counting the support score equation (2) was used. The item set candidate was combined into antecedent A and consequent B. For example in counting support score in Music and Game item, the support score from the total appearance of Music and Game occur simultaneously once for every visitor, that is; 4 divided by 10 (the total of visitors) equal for 0.4 or 40%. By determining the minimum support then the frequent item set (F_2) was {{News, Music}, {News, Game}, {News, Technology}, and {Music, Game}}.

Table 5 Candidate 3 – *Item set*

Combination of Article Category

A	B	Support
News, Music	Game	4/10 40%
News, Music	Technology	2/10 20%
News, Game	Technology	3/10 30%
Music, Game	Technology	2/10 20%

In item set candidate 3, the minimum support was defined as frequent item set or (F_3) for {News, Music, Game) as seen in table 5 above. For example, in counting the support score in {News, Music} and {Technology}, the support score was obtained from the total appearance of News, Music and Technology once for each

visitor, that is; 2 divided by 10 (the total of visitors) which equal for 0.2 or 20%. Since the total item from frequent item set was insufficient for item set candidate 4, then the iteration was stopped.

In the step of forming association rule, the candidate of association rule A ⇨ B was formed, thus it needed an item from frequent item set to be the antecedent item A and consequent item B. The association rule was formed from the association rule's candidate which fulfilled the minimum requirement for confidence. In counting the confidence score the equality of (3) was used.

Table 6. Association rule candidate of F2

Rule	Confidence
If Music, Game, and Technology, then News	2/2 100%
If News, Game, and Technology, then Music	2/3 66.6%
If News, Music, and Technology, then Game	2/2 100%
If News, Music, and Game, then Technology	2/4 50%

Table 6 above showed the association rule candidate was formed from F2. To count the confidence score in "If News, Game, and Technology, then Music" rule, then the confidence score from the total of antecedent items {News, Game, and Technology} and consequent item {Music} appeared once for every visitors, that is; 4 divided by 3 (total antecedent item) = 0.666 or 66.6%. If the minimum confidence was 90% then the association rule that could be formed were "If Music, Game, and Technology, then News" and "If News, Music, and Technology, then Game".

Table 7. Association Rule Candidate of F3

Rule	Confidence
If Music and Game, then News	4/4 100%

If News and Game, then Music	4/9 44.44444444444 4%
If News and Music, then Game	4/4 100%

The table 7 above showed the association rule candidate was formed from F_3. For example, in counting the confidence score in "If Music and Game, then News", the confidence score from the total antecedent items {Music, Game} and consequent item {News} appeared once in every visitor was 4 divided by 4 (total of antecedent item) = 1 or 100%. By defining the minimum confidence then the association rule which could be formed were "If Music and Game, then News" and "If News and Music, then Game". If the visitor visited an article in News category then the association rule with confidence item will be saved. The last saved rule had its antecedent item as the recommended article category. Hence, the recommendation article for News category was took from the association rule formed by F2, "If Music and Game, then News". Then the recommendation article were Music and Game.

3.2. UML Design

Web design using Unified Model Language (UML) as a modelling system had its designing diagram ordered in a sequence below;

a). Activity Diagram

Activity diagram explained about the activity route in a planned system. The activity diagram in a developed website was the activity of visited article page which could be seen in Figure 2 below;
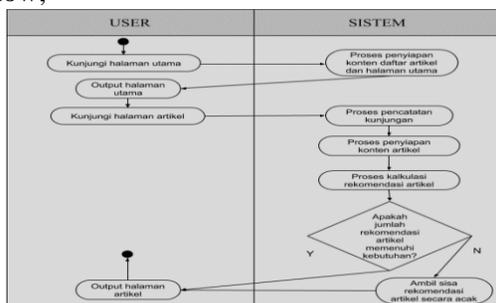

Figure 2. Activity Diagram on Visited Article Page

Figure 2 above showed the working route of activity diagram on visited article page. The route was started from the main page of the website. Then visited on the article page through the link in the main page. Then the system received the data identity of visitors and visited articles. With the data, the system was checked whether the article is available or not. If the article is available, the system will continue the process and if it is unavailable the user or visitor will be directed to the main page. The system will note and calculate the article recommendation from the collected data. In calculation process, Apriori Algorithm was used to find the frequent item set and the association rule form. Before showing the result, the system will check whether the total amount of article recommendation had fulfilled the necessity or not. If it had not fulfilled the necessity, then random article will be taken randomly for complete the deficiency.

b.) Sequence Diagram

Sequence Diagram was a diagram which described about the communication between available objects within the developed website. This diagram showed the interaction sequence done by the objects within the website.
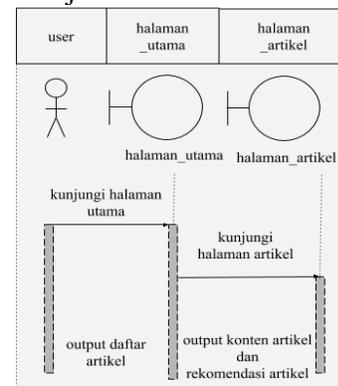

Figure 3. *Sequence Diagram* of Visited Article

Sequence diagram which showed in Figure 3 above was the article recommendation process. The steps was started from user visited the main page and then visited the article page. After that, the system will note down the visited article's

history and calculate the article recommendation. In the process of article recommendation calculation, Apriori Algorithm was used in finding the frequent item set and in forming the association rule. After that, user will get article page contained with article recommendation.

3.3. Designing Database

Database is a group of data consisting of connected table. Database had function to accommodate some table and query used as data source management. There are some ways to make and access the database, but in this study, mysql database was used. Mysql was used since it could help the database display based on designer's desire. The designed database table of database u217585741_mydb could be seen below;

Figure 4. Art7wing's Article Recommendation Database Relation

3.4. Design Interface

Website interface was arranged to facilitate the operational and to make simple and efficient interaction in catching visitor's attention. Also to make balance between technical function and visual element in order to be functioned as it is for.

3.5. Material Collection

This step was the step in collecting material related to the website developing process, either in the form of picture, text, or other materials from internet or self-made using supporting software.

3.6. Assembly

Assembly was the step in assembling the materials used for website development in order to create a website with mechanical component of Apriori Algorithm to give article recommendation. The programming language

used in implementing the Apriori Algorithm was PHP. The programming language was also used as the data to process the visitor history to be article recommendation.

3.7. Distribution

After the web application was done, then the distribution process was started. The distribution process was uploading the web application to virtual server. The website storage was in the form of file and folder with variety of access right to keep the website safety and development

3.8. Apriori algorithm on Article Recommendation

Apriori Algorithm on developed website was aimed to obtain article recommendation. The application of Apriori Algorithm could be seen through article page and matching the article recommendation result with Apriori Algorithm analysis result.

Figure 5. Article Recommendation on Article Page of News Category

Figure 5 above showed the recommended article if user visited the article on news category. It was seen that the item category of article recommendation showed were Music, Linux, and Computer. After that, matching the article recommendation result with Apriori Algorithm analysis result.

Table 8. Association Rule Candidate of F4

Association Rule Candidate	Confidence
If Music, Linux, and Computer, then News	31/31 100%
If News, Linux, and Computer, then Music	31/37 83.783783783784%
If News, Music, and Computer, then Linux	31/54 57.407407407407%
If News, Music, and Linux, then Computer	31/48 64.583333333333%

By looking at the Apriori Algorithm analysis result, it could be found that the item category on article recommendation was taken from the association rule antecedent "If Music, Linux, and Computer, then News" formed from the association rule candidate of F_4 as seen in table 8 above. The association rule used was the last saved association rule which had similar consequent score with visited article category. The item category of an article was taken randomly which published for 3 months based on the article category on the chosen antecedent association rule.

V. CONCLUSION

In conclusion, the association rule of "If Music, Linux, and Computer, then News" with confidence score 100% formed by Apriori Algorithm was used to give article recommendation on article page in News category. The visitor in article page could affect the association rule formed by Apriori Algorithm. In the analysis of higher frequency, the category of article item from visitor history could be used to find for the frequent item set. Then in forming the association rule, the forming of association rule of frequent item set was done. The association rule used for article recommendation was the last saved association rule which has similar consequent score with the visited article category. Therefore, the article recommendation was showed successfully in the article page of Art7wing.

For further research, the development on defining the minimum support and confidence needed to be done automatically by managing data from the previous result periodically, thus the association rule will be more accurate.

VI. REFERENCES

[1] Binanto. 2013. Perbandingan Metode Pengembangan Perangkat Lunak Multimedia. *Seminar Nasional RITEKTRA*. 153-160.

[2] Fadlina. 2014. Data Mining untuk Analisa Tingkat Kejahatan Jalanan dengan Algoritma Association Rule Metode Apriori. *Jurnal Ilmiah Informasidan Tekhnologi Ilmiah*, Vol. 3, No.1.

[3] Kusrini and Luthfi, E.T. 2013. *Algoritma Data Mining*, Yogyakarta: ANDI.

[4] Sugiyono. 2016. *Metode Penelitian Kuantitatif, Kualitatif dan R&D*, Bandung: PT Alfabet.

[5] Sutopo, A. H. 2012. *Teknologi Informasi dan Komunikasi dalam Pendidikan*, Yogyakarta: Graha Ilmu.

[6] Trimarsiah, Y., and Arafat. 2017. Analisis Dan Perancangan Website Sebagai Sarana Informasi Pada Lembaga Bahasa Kewirausahaan dan Komputer Akmi Baturaja, *Jurnal Ilmiah MATRIK*, Vol. 19, No.1, 1-10.