

Recommendation System for Property Search Using Content Based Filtering Method

Tessy Badriyah, Sefryan Azvy, Wiratmoko Yuwono, Iwan Syarif

Informatics and Computer Engineering Department

Electronic Engineering Polytechnic Institute of Surabaya (EEPIS)

Kampus Politeknik Elektronika Negeri Surabaya (PENS) Keputih Sukolilo Surabaya 60111, INDONESIA

*Corresponding author: tessy@pens.ac.id

Abstract— Development of technology causes many business industries to migrate from offline business systems to the e-commerce world. One of the most popular e-commerce frequented by potential buyers is the property site. Considering that the property is one of the essential requirements for living, and furthermore it is also one of the most prized assets one can have. In this research, we develop a web-based recommendation system in choosing a property using content-based filtering method. The recommendation system provides property information based on user behavior by searching advertising content previously searched by the user. Each time the user selects the contents of the ad to display, this information will be stored into the database to be processed further in order to provide a recommendation. The application system will present the same product recommendation, in accordance to the profile / criteria and preference of the prospective buyer. Therefore, the recommendation system will assist prospective buyers in determining the choice of property product they want to buy, and this process can be provided by the recommendation system in a short time.

Keywords—e-commerce, recommendation systems, content-based filtering

I. INTRODUCTION

The effect of internet on buyer's housing search has been examined by Leonard et al (2003). They proved that with the role of a broker will be able to reduce the search time of the property, otherwise by using Internet it can increase the intensity of the search but not reduce the search time. From here we conclude that it takes one facility on the Internet to buyer's housing search to reduce the search time by using a recommendation system. Recommendation system can give suggestion of property which is favored by user so that search time can be reduced.

While on the property sites that exist today, most of them only have an information provided by the category of houses, shop, office, boarding house. In this case information is presented without a learning process to provide a recommendation to the user. In addition to categories of property types, there are also categories based on the size, whether the property is sold or rented, the status of the ownership letter, the facility or the contents of the house and the location of the property by province and city. But there are two types of users who search for information. The first user is one who already knows the wants and the specifics. While the second user is a person who does not know their needs

specifically. The latter needs to be noticed, given the recommendation facility so that the user knows the required property.

This research attempts to accommodate the needs of the second type of user as described in the second paragraph by providing a recommendation system on the property site. Recommendations are made using the Content Based Filtering (CBF) method that works by comparing the detail of the property of the advertised property product with the advertisements that are often seen by users. In the other words, the CBF method looks for the similarities of recommended advertisements content like home or shop category, property status sold or rented, house size and others according to the previous advertisements viewed by the user. Systems using the CBF method should record the details of the advertisements viewed by the user so that the basic data are formulated for analysis. Each recorded data is assigned a weighted value based on the frequency of the data appearing in the search process performed by the user.

The results of CBF method analysis provide advertisement recommendation that suitable with current user behavior and also majority of data tendency obtained from other user behaviors. This combines other user behavior, to obtain advertising data that is experiencing an increase in search or trend. It is hoped that with this recommendation system, it will make it easier for users to get properties that suit their needs and increase property sales.

The purpose of this research is to build a recommendation system on property sites that has the following contributions:

1. This research uses content-based filtering by grouping properties by region, category, ownership status, property letter status, property size, content or facility property
2. The recommendation system provides an advice and property information based on advertising content previously searched by the user. After that, we use Apriori Algorithm that will make the formation of search database based on the advertisement content by analyzing the weight of data based on the frequency of view on the user's search.

The rest of the paper is organized as follows. The next section provides a brief overview of some related research work related to recommendation systems. Section 3 explains about the algorithm that underlies the recommendation process

and also the design of recommendation system framework. Section 4 explains in detail how the method we propose gives recommendations to users. Section 5 describes results of experiments and discussion of the results. The final section provides conclusion.

II. RELATED WORKS

Recommendation system can be distinguished into three method: (1) Collaborative filtering method that compare active user with preference of other users, while in (2) Content-based Filtering depends on active user's choices and a combination of both called the (3) Hybrid filtering.

Recommendation systems have been widely used in e-commerce primarily on the sale of goods. Before discussing the research related to the recommendation system on the property site, we will discuss the research that developed the e-commerce recommendation system to sell the goods because the most common recommendation system used here.

In the following, we will discuss one of the methods in content based filtering related to the method used in our research: the association rule method. Rolfnes et al (2016) made improvements to the recommendations by using aggregated association rules. While Jomsri (2014) uses association rules based on user profiles to provide book recommendation information for digital library. Other researchers such as Chen et al (2014) uses the association rules mining used for small online retailers. And the other, Kim et al (2013) builds a decision making model that can identify and provide proper exercise recommendations.

To make things better, some researchers combine association rule with other methods on the recommendation system. Such as research conducted by Badriyah et al (2017) who build a recommendation system on e-commerce by combining product descriptions on content-based filtering. In this study, it was found that merging the two methods can give satisfactory results on the performance analysis. Other research conducted by Tewari and Priyanka (2014) works on a book recommendation system by combining collaborative filtering and association rule mining methods.

In contrast to the recommendation system that is widely used in e-commerce for the sale of goods, there are only a few studies conducted for the recommendation system on property sites. The following review will describe some of the research related to the recommendation system on the site property, as well as the relevance of the research we will undertake

Yuan *et al.*, (2013) developed user oriented recommendation system for property site based on user behaviours. From there then the author used case-based reasoning and ontological structure method to generate recommendation of property given to the user. However, this study accommodates the representative user requirements and constraints which consist of location, housing unit property and the price obtained and the preference and status of the user profile. Comparing to Yuan et al., (2013), the difference with the previous paper by Bond et. al (2000) is the feature elements used. The difference between the two papers with our study is on the use of user behavior. Rather than focuses

on user behavior of requirements and constrained derived from user information, in this research we focus on the content of the properties that are viewed by the user.

In this research we use content based filtering method using algorithm apriori in association rule mining on property site. The recommendation is based on advertisements information that the user has previously searched for. Apriori Algorithm that will make the formation of search database based on the advertisement content by analyzing the weight of data based on the frequency of view on the user's search.

III. RECOMMENDATION SYSTEM DESIGN

This section will explain the algorithm underlying the recommendation process with content based filtering using Apriori algorithm. Then after that we discuss framework recommendation system designed in this research.

A. The Recommendation System

The recommendation system is a methodology that aims to assist users by providing recommendations to users when users are faced with large amounts of information. Recommendations are expected to help users in the decision-making process, such as what items to buy, what books to read, or what music will be heard, and others (Hoga Saragih, 2013).

A personalized recommendation system must be familiar to every existing user. Each recommendation system should build and maintain user models or user profiles of user interest. For example, the recommendation system on the Amazon website stores every customer purchase transaction, customer comments, and reviews / ratings given by customers to a product

There are two approaches that can be done to form a user profile, namely implicit and explicit approach. The implicit approach stores and learns user behavior within the system to build user profiles. User behavior can be likes / dislikes, ratings, and so on. Explicit approach namely asks the user directly the description of the item favored by the user.

1) Content Based Filtering

Content-based Recommendation System uses content availability (often called features, attributes or characteristics) of an item as a basis for recommendation (Hoga Saragih, 2013). For example, a movie has content such as genre, author, release year, etc., or a document file has text content contained within it.

Content-based filtering provides a recommendation based on the analysis of similarity items that have been assessed by users. Content-Based Filtering shapes its user profile based on the item-forming attributes. For example, for a document, its formatting attribute is the word that is contained in the document. This user profile generator parameter is also assigned a weighted value based on certain criteria. The algorithm steps are as follows:

a. One product is divided according to a vector of its constituent components.

b. The system creates a user profile based on the vector weights of the components that make up an item. User profile creation can use the TF-IDF (term frequency-inverse document frequency) algorithm. TF is the number of terms in a document. While the IDF value can be calculated using the formula:

$$\text{idf}_i = \log\left(\frac{n}{\text{df}_i}\right) \quad (1)$$

Where:

n is the sum of all documents while df is the number of documents having term i .

Based on that user profile, the system will estimate the likes or dislikes of an item based on the analysis of the user profile resemblance with the vector component of the item shaper. If the system estimates that the item will be favored by the user, then the item will be recommended to the user.

Content-based recommendation systems have several advantages:

- a) Content-based recommendation systems can explain how recommendations are obtained.
- b) Content-based recommendation systems may recommend items that have not even been rated by anyone.

However, content-based recommendation systems also have some disadvantages as follows:

- a) Content-based recommendation system does not have the ability to deliver unexpected recommendation results (Serendipity Problem).
- b) Content-based recommendation system requires a user profile that contains user preferences. For new users who have not done any activity and do not have enough user profiles, the recommendation system cannot provide a reliable recommendation to them (Cold Start Problem).

2) Association Rule Mining (Apriori)

Apriori algorithm includes the type of association rules on data mining. This rule states the association between some attributes is often called affinity analysis or market basket analysis (Hoga Saragih, 2013). Association analysis or association rule mining is a data mining technique for finding associative rules between a combinations of items. Associated analysis is also known as one of the data mining techniques that form the basis of various other data mining techniques. Particularly one of the stages of an association analysis called the analysis of frequent pattern mining attracts the attention of many researchers to produce an efficient algorithm.

The importance of an associative rule can be known by two parameters, support (the value of support) is the percentage of the combination of items in the database and confidence (value of certainty) is the strength of the relationship between items in associative rules. The basic methodology of association analysis is divided into two stages: frequent itemsets and generate an association rules that will be explained in the following.

Frequent Itemsets

This stage looks for a combination of items that meet the minimum requirements of the support value in the database. The value of an item's support is obtained by the following formula:

Support (A) = Number of Transactions containing A / Total Transactions

While the support value of 2 (two) items is obtained from the following formula:

Support (A \cap B) = Number of Transactions containing A and B / Total Transactions

Generate an Association Rules

After all of the frequent itemsets was found, we obtained the associative rule that meets the minimum requirement for confidence by calculating the confidence of the associative rule A and B. The confidence value of rule A and B was obtained from the following formula:

Confidence = P (B | A) = Number of Transactions containing A and B / Number of Transactions containing A

B. The Recommendation System Framework

The work process recommendation system workflow proposed in this study can be seen in Figure 1 below:

Framework Recommendation System Design

1. The user visits several advertisementss that exist on the system, after that the ad data visited by the user is stored on the system database
2. System conducts collection of words visited by users based on: Title, Description, Address and Ad Description.
3. The system then calculates the weight of selected words using the TF-IDF method
4. The word with the highest TF-IDF value will be found in the ID of the property ad on the database containing each of those words.
5. The ID of displayed advertisements will be the itemset to be analyzed using the Apriori method.
6. From the Itemset, the formation of Frequent Itemset is the Itemset that appears most frequently.
7. The results obtained from the Frequent Itemset are then used to form the Association Rule pattern
8. The advertisement recommendation data is then displayed based on previously formed association rule patterns as parameters.

Fig. 1. Recommendation System Framework

IV. EXPERIMENTS AND ANALYSIS

In this section we will describe the experiments conducted under the design of Recommendation System Framework.

A. Storing data based on user behaviour

Experiments begin with users who visit some of the property advertisements they are interested in, after that the system will store the product information of this property as a base to find the recommended property data for the user. From Figure 2, it can be seen that the user visited 2 property product data with ad id 38 and 40 as shown in the array display below:

[advertisement_id] => 38
[advertisement_name] => Rumah Dijual Rungkut Mapan Tengah Surabaya
[description] => HARGA : 2 M NEGO, SHM, IMB
LT : 189 M2 (10.5X18)
LB : 215 M2 (1.5LT)
KT : 3 , KM : 3
HADAP : UTARA
ROW JL. : 2 MOBIL
LISTRIK : 4400 W
FACILITIES : KOSONGAN, AC 2 UNIT, DEKAT DENGAN MC DONALD, GIANT.
FOR FURTHER INFO, YOU CAN CONTACT EVAN N
[address] => RUNGKUT MAPAN TENGAH
[district] => RUNGKUT
[advertisement_id] => 40
[advertisement_name] => Rumah Dijual Royal Residence Balmoral Surabaya
[description] => HARGA MASIH NEGO, SEMI FURNISH
[address] => ROYAL RESIDENCE BALMORAL
[district] => WIYUNG

Fig. 2. Storing data based on user's behaviour

B. Calculation Weight of Words Using TF-IDF Method

These advertisements are then collected and searched for the combined weight of those advertisements using the TF-IDF method, and from the most frequently generated words along with their TF-IDF values, we obtain the top 10 words, which will be used as itemsets to be analyzed using apriori as shown in Table I.

C. Generate an Association Rules

The next step is to make the pattern of association rule of the frequent itemset along with its confidence value that has been obtained as follows:

38=100%
 39=100%
 34 => 38 = 100%
 54 => 53 = 100%
 54 => 40 = 100%
 54 => 40,53 = 100%
 54 => 39 = 100%
 54 => 39,53 = 100%
 54 => 39,40 = 100%
 54 => 39,40,53 = 100%
 53 => 54 = 100%
 53 => 40 = 100%
 53 => 40,54 = 100%
 53 => 39 = 100%
 53 => 39,54 = 100%
 53 => 39,40 = 100%
 53 => 39,40,54 = 100%
 40,53 => 54 = 100%
 40,53 => 39 = 100%
 40,53 => 39,54 = 100%

TABLE I. THE TF-IDF VALUE OF THE ADVERTISEMENT KEYWORD THE USER VISITED

WORD	TF-IDF VALUE
wiyung	0.301029996
w	0.301029996
utara	0.301029996
unit	0.301029996
tengah	0.301029996
survey	0.301029996
shm	0.301029996
semi	0.301029996
rungkut	0.301029996
royal	0.301029996
row	0.301029996
mobil	0.301029996
mc	0.301029996
residence	0.301029996

It can be seen in the previous data, the pattern that is formed, among others:

- If a user visits Advertisement with id 34, he will visit id 38 with 100% confidence value,
 - If a user visits Advertisement with id 54, he will visit id 53 with 100% confidence value,
 - If a user visits Advertisement with id 54, he will visit id 40 with 100% confidence value,
- And so on until the pattern is no longer formed.

D. Recommendation of properties

Finally, the system gives the recommendation data of property product to user based on preference of previously visited advertisement data where the pattern of association rule above will be used as reference. The next process is to search the property recommendation data in the database as shown in Figure 3 below, where based on advertisement preferences visited by users for example, advertisements with id 38 and 40, will provide the recommendation advertisements with id 39, 40, 36, 38, 53, and 54 as shown in the picture below

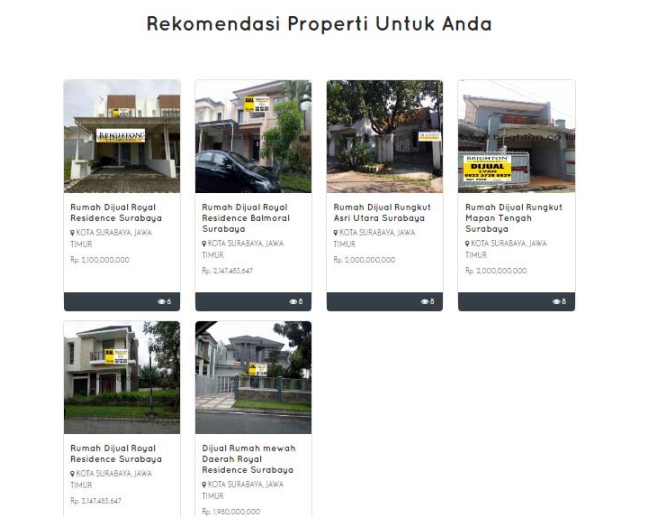


Fig. 3. Recommendation Page

V. CONCLUSIONS

This research develops Recommendation System for property search using Content Based Filtering method. Based on the design stage until the testing conducted, it can be concluded that the system created has been successful in giving recommendation of property advertisement based on behavioral data of system usage. The value of support and confidence is a value that affects many of the least recommended product recommendations. The greater the value of support and confidence given, fewer recommendations will be displayed. The smaller the value of support and confidence given, the more recommendations are displayed. The experiment results show that the system is able to provide recommendations with more than one keyword user preference.

REFERENCES

- [1] T. Badriyah, E.T. Wijayanto, I. Syarif, P. Kristalina "A hybrid recommendation system for E-commerce based on product description and user profile", 2017 Seventh International Conference on Innovative Computing Technology (INTECH), 2017, 16-18 Aug. 2017. p.95-100.
- [2] T. Rolfnes, L. Moonen, S. D. Alesio, R. Behjati and D. Binkley, "Improving Change Recommendation using Aggregated Association Rules," 2016 IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR), Austin, TX, 2016, pp. 73-84.
- [3] A. S. Tewari and K. Priyanka, "Book recommendation system based on collaborative filtering and association rule mining for college students," 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, 2014, pp. 135-138.
- [4] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014), Luton, 2014, pp. 130-134.
- [5] J. Chen, C. Miller and G. G. Dagher, "Product recommendation system for small online retailers using association rules mining," Proceedings of the 2014 International Conference on Innovative Design and Manufacturing (ICIDM), Montreal, QC, 2014, pp. 71-77.
- [6] A. S. Tewari, A. Kumar and A. G. Barman, "Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining," 2014 IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 500-503.
- [7] F. Min and W. Zhu, "Mining top-k granular association rules for recommendation," 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), Edmonton, AB, 2013, pp. 1372-1376.
- [8] J. K. Kim, K. H. Lee, D. K. Park and E. Y. Jung, "Context-Aware U-Health Service: Identification of Exercise Recommendation Factors and Creation of Decision-Making Model Using Association Rule," 2013 International Conference on Information Science and Applications (ICISA), Suwon, 2013, pp. 1-4.
- [9] X. Yuan, JH. Lee, SJ. Kim, YH. Kim, "Toward a user-oriented recommendation system for real estate websites". Information Systems, v. 38, n. 2, p. 231-243, 2013/04/01/ 2013. ISSN 0306-4379.
- [10] M.T. Bond, M.J. Seiler, V.L. Seiler, B. Blake, "Uses of websites for effective real estate marketing", Journal of Real Estate Portfolio Management 6(2) (2000) 203-210.
- [11] V.Z. Leonard, H.J. Ken, I.A. Randy, "Internet use and real estate brokerage market intermediation", Journal of Housing Economics 12 (2003) 134-150.
- [12] J.S.Ford, R.C.Rutherford, A.Yavas, "The effects of the Internet on marketing residential real estate", Journal of Housing Economics 14 (2005) 92-108.