International Journal of Computer Engineering and Information Technology

VOL. 10, NO. 1, January 2018, 1–5
Available online at: www.ijceit.org
E-ISSN 2412-8856 (Online)

Recommendations System for Purchase of Cosmetics Using Content-Based Filtering

Joanna Cristy Patty[1], Elika Thea Kirana[2] and Made Sandra Diamond Khrismayanti Giri[3]

[1, 2, 3] Magister Teknik Informatika, Universitas Atma Jaya Yogyakarta, Yogyakarta, 55281, Indonesia

[1]cristyjoanna18@gmail.com, [2]elikatheaa@gmail.com, [3]madesandragiri@gmail.com

ABSTRACT

Over the years, technology has been growing fast. Currently, the rapid development of this technology also has affected the rapid development of the cosmetics industry. Progress in the beauty industry in Indonesia shows many improvements. In Indonesia, there are various specifications of the cosmetic market. Many cosmetic specifications make prospective buyers find it difficult to choose the type of cosmetics that suit their needs. Prospective buyers usually come to the cosmetics counter to find information about cosmetics as desired. However, not all counters can provide appropriate recommendations for the prospective buyer that have no intention of different specifications. The way it is less effective and it also takes a long time.

The solution to solve the problem is to make the design of the recommendation system for the selection of cosmetic purchases. The recommendation system aims to facilitate the customers take the right decision in determining which products to be used. The method used in the system of recommendations have been made is Content-Based Filtering method. Content-Based Filtering is done based on common approaches item profiles and user profiles.

The system design is built on cosmetics to provide recommendations in accordance with the required specifications of users. Specs shown include cosmetic type, skin type, usage, price, description and pictures..

Keywords: *Content Based Filtering, Recommendations System, Cosmetics.*

1. INTRODUCTION

In this globalization era many needs associated with technology. Many of the latest technologies that have sprung up along with the times and the need for technologies that increase. Information technology today is very helpful in supporting human life almost all aspects of life. Technological developments such as the internet (Maharani & Gunawan, 2015) make it easier for people to search for information. Easy and quick information obtained will be able to make people more and more knowledge. Current technology is developing very rapidly. The rapid development of this technology also has affected the rapid development of the cosmetics industry today. Progress in the beauty industry in Indonesia at this time to show improvement. Based on data from the Ministry of Industry (2016), The industrial market growth averaged 9.67% per year in the last six years (2009-2015). It is estimated that of the market (market size) cosmetic market is Rp. 46.4 trillion in 2017's. With this amount, Indonesia is a potential market for entrepreneurs industry.

Sigma Research Indonesia conducts research to Indonesia by segments 1200 women aged 15-55 years. From these studies revealed that there are several factors into consideration women in purchasing cosmetic products. The biggest factor is the percentage formula matches the facial skin (79.4%), followed by long-lasting products (67.4%), lightweight formula (62.2%), as well as color selection, kosher, with percentage price respectively above 50%.

According Wahyo U & Anggriawan, (2015) a recommendation system is a model application of the results of observations on the circumstances and wishes of the customer. Therefore, a recommendation system requires a model of precise recommendations that were recommended in accordance with the wishes of customers, as well as enabling customers to take the right decision in determining which products will be used.

Of the above problems, the authors of this research will make a recommendation system design cosmetics using Content- Based Filtering method using TF-IDF algorithm. Content-Based Filtering is done based on common approaches item profiles and user profiles (the Goddess, 2013). TF-IDF algorithm is a way to give weight to a category or term relationship in a document. The system is expected to help someone who is going to buy cosmetics to conform to the required specifications.

2. LITERATURE REVIEW

As the technology advances, more and more are also easily available information to solve a problem that occurs in everyday life. Applications that utilize a recommendation system already is growing rapidly at the moment, it can be seen from various fields such as education, business, automotive, agriculture, marketing and sales.

Recommendation systems are typically used to take a decision. According to the Son, et al., (2015) in his study of recommendations elective courses the students with Content-Based Filtering and Collaborative Filtering, a recommendation system will provide different recommendations to each user, not just provide a list of the items most in demand, but rather provide advice on items that may be appropriate for users. That is, each user will get different recommendations, according to the profile and user interest.

Maharani & Gunawan, (2015) made a research about the car selection based on Demographic and Content-Based Filtering. In this research combines methods Demographic and Content-Based Filtering in order to make recommendations that resulted really fit the profile and preferences will buyers. Demographic Filtering method is commonly used to handle the newly registered user or users who do not have the profile or records of transactions or activity in the system (history). In this study, recommendations are given for Content-Based Filtering is obtained based on the similarity between user preferences will car with specs of cars available in the system.

Subsequent research Arishintha (2013) made a recommendation system for the marketing of merchandise sales. In this study, using Content Based Filtering method based mobile android. The purpose of this study is to provide the best information for sales in marketing their products to shops or stores. There are two approaches to filtering information that is collaborative filtering and content-based filtering. On Collaborative Filtering recommendation process based on the similarity between users who opt into the system, this approach is divided into two, User-Based Collaborative Filtering and Item-Based Collaborative Filtering. Content Based Filtering on the recommendation of an item to a user based on the description of the item and the profile of the user's interest.

Subsequent research Goddess (2013) made a recommendation system for drug sales. In this study, using a method based mobile Content Based Filtering. The purpose of this study is to provide an application that can provide doctor's recommendation regarding the relevant information to be visited by MedRep (Medical Sales Representative) based approaches Content Based Filtering. Content Based Filtering is done based on common approaches item profiles and user profiles.

Subsequent research Joanna, et al (2017) made a recommendation to purchase the system design cosmetics. In this study, using Content Based Filtering. The purpose of this study is to establish a system of cosmetic recommendations for someone who will buy cosmetics that suit your needs according to his skin color and skin type. In this study, using Content-Based Filtering approach. Content Based Filtering is done based on common approaches item profiles and user profiles. Then measure the similarity of the item profiles and user profiles using the Cosine Similarity.

3. THEORETICAL BASIS

3.1 System Recommendations

A recommendation system is a system to provide and recommend an item to make a decision that is desired by the user (Ungkawa, et al., 2013). There are various purpose of implementing a recommendation system, including increasing the amount of sales of goods, can attract users, get a recommendation in accordance with the wishes. The accuracy of a system on one of them rely on the algorithm used. There are various methods of approach used to solve the problems in the system recommendation, among others Content-Based Filtering, and Hybrid Collaborative Filtering Recommender System.

3.2 Content-Based Filtering

Content-Based Filtering establish its profile by forming attributes of an item. Such systems are commonly used to recommend web pages, TV programs, news articles, etc. (Sharma & Gera, 2013). According Nastiti (2013), using content-based filtering approach is different from collaborative filtering to select and rank the items to the user based on the similarity of users to other users who like similar items in the past. Content-based filtering approach is to select and rank the items based on similarity perform user profile and profile items. Content-Based Filtering algorithm can be described in the following steps:

1. An item divided by its constituent component vectors.
2. The system will create a user profile based on weight vectors forming part of an item. Users can use the profile creation algorithm TF-IDF (term frequency-inverse document frequency). TF is the number of terms in a document. While the IDF can be calculated using the formula:

$$idf = log_2 \frac{n}{df} \quad (\mathbf{3.1})$$

where n is the number of documents, whereas df is the number of documents that have the term i.

3. Based on the user profile, the system will estimate the ratings of like or dislike of an item based on the user profile similarity analysis with item-forming component vectors. If the system estimates that the item will be favored by the users then the item will be recommended to the user.

3.3 Cosine Similarity

According to Adi (2010) calculation of similarity between documents is done by calculating the Cosine Similarity between vectors and vector querying document collections. Here is the equation of the algorithm method Cosine Similarity:

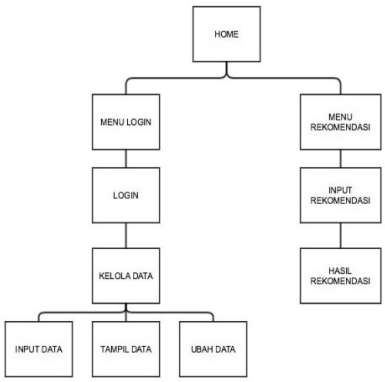
$$Similarity(x,y) = \frac{\sum_{i=1}^{t} x_i y_i}{\sqrt{\sum_{i=1}^{t} x^2{}_i \cdot \sum_{i=1}^{t} y^2{}_i}} \qquad (3.2)$$

Where:
x and y are distinct vectors.
xi = i existing term on the vector x
yi = i existing term on the vector y
Cosine Similarity of the calculation result, the highest calculation result will be a vector which is the nearest and the vector you want to compare.

4. RESEARCH RESULT

The design of the system is part of a software development methodology that is done by going through the stages of analysis, design meant to illustrate in detail how an application to be built. At the design stage of this system as well, as a continuation of the analysis system, which at the time described the draft system design to be built prior to encoding into a programming language that will be built by a programmer. In designing a system can't be separated from the results of the new analysis can be made of a draft system.
Recommended System Selection Purchasing cosmetics is a system built to provide cosmetic recommendation in accordance with the requirements desired by the user. Recommendations are displayed based on the specifications entered by the user. The method used in determining the recommendation is Content-Based Filtering. Calculation method on Content-Based Filtering through two stages. First calculate the value of the term in a document using TF-IDF algorithm. The second is to calculate the similarity between the user's profile to the profile items using algorithms Cosine Similarity

4.1 Architecture Design



4.2 Data Needs Analysis

Data needed in Planning Preparation electoral system on cosmetics purchases are:

Table 4.1. Category Type Cosmetics

No.	Cosmetics_Type
1	Facial Wash (Wash face)
2	Toner
	Moisture
4	Sun protection
5	Powder
6	Mask
7	Serum

Table 4.1. Categories of Use

NO	Use
1	Morning
2	Night

Fig. 4.2. Categories Skin Types

source : www.google.com
Table 4.2 Category Price

No.	Price
1	Rp. 10.000 - Rp. 50,000
2	Rp. 50.000 - Rp. 100,000
3	Rp. 100,000 - Rp. 150.000

4	Rp. 150000-200000
5	> Rp. 200,000

4.3 Calculation Of Recommendations

Calculation of the recommendations on the system is using Content- Based Filtering with TF IDF algorithm and using a calculation the degree of similarity between the TF-IDF cosmetics of data residing on the database and the data TF- IDF cosmetic use Cosine Similarity user input. TF is the number of terms in a document, in terms of this study were divided into 17 categories. Table category to be related to the data tables cosmetics. Cosmetics category can be seen in Table 4.1 - 4.1- 4.2 and Figure 4.2 on page 7 - 8. Number each table shows the id of each category.

Calculating system TF (Term Frequency) every category on all cosmetics. TF category of every cosmetic is expressed in the form of id each category can be seen in Table 4.3

No.	Cosmetics Name
1	Pand's Pure white
2	Pand's smooth pores
3	Pand's white beauty
4	Pand's oil control
5	Pand's no blackhead
-	----------
36	UV Tint
37	spot Care
38	Cream
39	compact powder
40	Oil Control Mask

Unknown user enters the desired cosmetic specifications. Can be seen in the following Table:

Category	input
type Cosmetics	1
Skin type	3
Use	1
Price	1

Next, calculate the inverse document frequency (IDF) that belongs to each category according to the formula:

$$idf = log_2 \frac{n}{df}$$

n is the number of cosmetic Cosmetics whereas df is the number that has a term i. Here is an example idf value for the category of cosmetic type, namely Facial Wash (Face Wash)

$$idf_{facial\ wash} = log_2\left(\frac{40}{9}\right)$$

$$idf_{facial\ wash} = 2{,}15200$$

After getting the IDF value of each category and the TF value of each cosmetics, the system calculates the length of each cosmetics can be calculated by the equation

$$\boldsymbol{length} = \sqrt{\begin{matrix}(\boldsymbol{idf\ Cosmetics\_Type})^2 + \\ (\boldsymbol{idf\ Usage})^2 + \\ (\boldsymbol{idf\ Skin\_Type})^2 + (\boldsymbol{idf\ Price}\ )^2\end{matrix}}$$

where :
idf_Cosmetics_Type are idf_Cosmetics_Type that have a term i
idf_Usage is idf_Usage which has a term i
idf_Skin_Type is idf_Skin_Type that has a term i
idf_Price is idf_Price which has a term i
The data used to calculate the length of cosmetics is the idf value of each category. Here is an example idf value used to calculate length Cosmetics Pand's Pure white can be seen in the table

$$length = \sqrt{\begin{matrix}(4{,}361)^2 + (0{,}1352)^2 + \\ (0.1036)^2 + (0{,}0770)^2\end{matrix}}$$

$$\boldsymbol{length = 2.2241}$$

After getting a TF-IDF and length every cosmetic and users, so would be the closeness between cosmetics with user input. From the results of these calculations, we can see the value of the highest closeness that would be my recommendation to the users of cosmetics. The formula used is the cosine similarity.

$$similarity(x, y) = \frac{x_a y_{a+}\ x_b y_{b+}\ x_c y_{c+}\ x_d y_d}{\sqrt{(x_a^2 x_b^2 x_c^2 x_d^2)(y_a^2 y_b^2 y_c^2 y_d^2)}}$$

$$similarity(x,y) = \frac{\begin{matrix}(4{,}361\ .4{,}361) + (0{,}1352\ .0{,}1352) + \\ (0{,}1036\ .0{,}1306) + (0{,}0770)\end{matrix}}{\sqrt{\begin{matrix}(4{,}361^2 + 0{,}1352^2 + 0{,}1036^2 + 0{,}0770^2) * \\ (4{,}361^2 + 0{,}1352^2 + 0{,}1036^2 + 0{,}0770^2)\end{matrix}}}$$

$$similarity(x, y) = 0.9422$$

Table similarity calculation results

No.	Cosmetics Name	cosine Similarity
1	Pand's Pure white	**0.942247345**
2	Pand's smooth pores	**0.831156412**
3	Pand's white beauty	**0.901293932**
4	Pand's oil control	**0.959050418**

5	Pand's no blackhead	**0.916603302**
-	----------	
36	UV Tint	**0.077672175**
37	spot Care	**0.146823015**
38	Cream	**0.159296268**
39	compact powder	**0.077672175**
40	Oil Control Mask	**0.062376364**

From the above calculation, the 3rd highest cosine similarity value is a cosmetic that would be recommended. Cosmetics with the highest similarity score is Pand's oil control, Pand's Pure white, and Pand's no bleakhead. The results of manual calculations above are matched with the recommendation of the planning application recommendations.

5. CONCLUSION

Based on the results of the discussion in the previous chapters, it can be concluded is that:

1. Development Planning Recommendation System Selection Purchasing cosmetics using methods *Content-Based Filtering* cosmetic provide recommendations to users according to the user's desired specifications.
2. Method *Content-Based Filtering* successfully implemented using algorithms TF-IDF as a comparison of data existing cosmetic cosmetics of data in the database with the user input.

Development Planning Recommendation System Selection Purchasing this Cosmetics has succeeded in making calculations using Content-Based Filtering method for recommending a cosmetic that will be used by developers.

REFERENCES

[1] Dewi, LS, 2013. System sales on drug use content-based filtering approach based mobile android. Pp. 1-202.
[2] Maharani, H. & Gunawan, FA, 2015. Car Recommendation System Based on Demographic and Content-Based Filtering. Telematics Journal, IX (2), pp. 64-68.
[3] Nastiti, P., 2013. RECOMMENDATION SYSTEM OF AGRICULTURAL CROPS USING CONTENT based filtering. pp. 1-141.
[4] Patty Joanna, et al., 2017 ELECTORAL SYSTEM DESIGN RECOMMENDATIONS FOR PURCHASE OF COSMETICS USING CONTENT-based filtering Pp. 1-7.
[5] Son, AFH, Mahmudy, WF & Setiawan, BD, 2015. Recommendation System Electives Students with Content-Based Filtering and Collaborative Filtering (Case Study: the UB). Repository PTIIK UB's Student Journal, V (1), pp. 1-11.
[6] Sharma, L. & Gera, A., 2013. A Survey of Recommendation System: Research Challenges. International Journal of Engineering Trends and Technology (IJETT), IV (5), pp. 1989-1992.
[7] The Ministry, 2016 [online] available at: http://sigmaresearch.co.id/tren-dan-perilaku-pasar-kosmetik-indonesia-tahun-2017/ accessible pada: August 2, 2017
[8] Ungkawa, U., Rosmala, D. & Aryanti, F., 2013. APPLICATION DEVELOPMENT WITH TRAVEL CASE BASE recommender reasoning. Informatics Journal, II (4-3), pp. 1-9.
[9] Wahyo U, BT & Anggriawan, AW, 2015. Recommendation System Package Se Malang Hybrid Method Using Content-Based and Collaborative. Scientific Journal of Information Technology and ASIA, IX (1), pp. 6-13.