

PAPER • OPEN ACCESS

Item Clustering as An Input for Skin Care Product Recommended System using Content Based Filtering

To cite this article: Villia Putriany *et al* 2019 *J. Phys.: Conf. Ser.* **1196** 012004

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

Item Clustering as An Input for Skin Care Product Recommended System using Content Based Filtering

Villia Putriany¹, Jaidan Jauhari², Rahmat Izwan Heroza³

^{1,2,3} Faculty of Computer Science, Sriwijaya University, Indralaya, Indonesia

Corresponding author: rahmatheroza@unsri.ac.id (Rahmat Izwan Heroza)

Abstract. A lot of skin care products sold in the market shows us that skin care is an important part of lifestyle for both women and men. Considering that everyone has different skin profile, a recommendation system is required to help and give a personalised suggestion products based on the user's preferences. Recommendation system provides suggestions that efficiently narrowing down the amount of informations so users will directed to the items which is most suitable for their skin. The method used in this paper is Content Based Filtering with K-means clustering for suggestion product calculation. The result is the system that can recommend skin care products for the users based on products they like, thus the users would find products those possibly suitable for their skin.

1. Introduction

Skin care has become one of the important parts of lifestyle not only for women but also for men. For example, the Americans spend more money on beauty products than on education[1] while skin care is the beauty products that has growing selling point and will continue every year[2]. It was proved by data of Ministry of Industry showing in 2015 domestic selling for cosmetic products reach Rp 11 billion with two categories growing fast and one of them is skin care products[3]. We can conclude that there are so many skin care products in the market and so many consumer spend money on that. Moreover, the dermatologists advice to use at least five steps for a skin care routine and this five steps consist of five different products. American Society for Dermatologic Surgery Association suggests to use product that suitable for the consumer skin type[4]. There is a high chance of taking the wrong products that will give unwanted reactions or even damage the skin and this possibility need to be minimized.

To avoid unwanted results by using unsuitable skin care products, a recommended system is required to guide consumer to choose products that might be suitable for their skin. Recommended Systems are software tools and techniques to provide suggestions of items to be use by a user. The suggestions relate to various decision-making processes, such as what video to watch, what music to listen, what items to buy or what online news to read. "Item" is the general term that is used to denote what the system recommends to users. Normally recommended system focuses on a specific type of item (e.g., CDs or news). Hence, its design, user interface and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item[5]. Recommended systems are typically categorized into collaborative filtering and content based filtering.

Collaborative filtering gives recommendation based on similarity of the users past ratings. This method is used widely, but the main limitation of collaborative filtering is that it is vulnerable to fraud or profile-injection attacks where a number of fake user profiles are inserted into the system to influence the recommendations that is made to the users which might have significant negative effects to the systems. Collaborative filtering performance degrades as the number of users and items increase. Because of this, collaborative filtering algorithm must be regularly updated over time.

Content based filtering gives recommendation for suitable items to users based on the descriptions of the items and user preferences. Content based filtering works for active users because they can provide accurate information even if the information from the other users is not sufficiently large. On the other side, content based filtering cannot give suitable recommendation if the content analyzed for an item is not contain convenient information for categorization. To address this problem, many studies tried to resolve this[6], one of them use clustering techniques which are k-means clustering and modularity clustering.

In this study, we use the content based filtering method to recommend skin care products because skin care products particularly have some direct relation. K-means clustering is used as clustering method because it produces good clustering outcome for many practical applications effectively [7]. And also, content based filtering could use items' features that are given by users working in web 2.0, such as tags, posts and opinions [8], including user experience about some product. Choosing skin care products should suitable for the consumer skin type and content based filtering could provide this recommendation because content based filtering matching up items descriptions with user preferences.

2. Research Methods

2.1. Content Based Filtering

Content-based recommendation systems try to recommend items similar to those a given user has chosen in the past. It assumes that user will rate items, with similar attributes, similarly[9]. The basic process performed by a content-based recommender consists in matching up the attributes of a user profile in which preferences and interests are stored with the attributes of a content object (item), in order to recommend new interesting items to the user. The recommendation process is performed in three steps, each of which is handled by a separate component:

- Content analyzer

When the information has no structure (e.g., text), some kind of pre-processing step is needed to extract structured relevant information. The main responsibility of the component is to represent the content of items coming from information sources in a form suitable for the next processing steps. Data items are analyzed by feature extraction techniques in order to shift item representation from the original form information to the target one. In this step we will do pre-processing to transform categorical data into vector so this representation is the input to the Profile learner and Filtering component;

- Profile learner

This module collects data representative of the user preferences and tries to generalize this data, in order to construct the user profile. Usually, the generalization strategy is realized through machine learning techniques, which are able to infer a model of user interests starting from items liked or disliked in the past. Machine learning techniques use in this research is k-means clustering because it improves the accuracy of user rating predictions and substantially increases the speed of online calculations;

- Filtering component

This step exploits user profile to suggest relevant items by matching the profile representation with the items to be recommended. The result is a binary or continuous relevance judgment (computed using some similarity metrics), the latter case resulting in a ranked list of potentially interesting items. In the above mentioned, the matching is realized by computing the cosine similarity between the prototype vector and the item vectors[6].

2.2. K-Means Clustering

One of the clustering algorithm is K-means, this algorithm aims to minimizing cluster performance index, square-error and error criterion. K-means received input data without class label. In K-means, computer will clustered input data unknowing the class target of the data. This kind of study is unsupervised learning. The input is data or object and k cluster. The algorithm will categorized data or object into k cluster. In every cluster there's center point or centroid that represent the cluster. First, choose some dots to represent the initial cluster focal points; second, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification. If the classification is unreasonable, we will modify it by recalculate each cluster focal points repetitively until we get a reasonable classification. K-means algorithm has advantages of briefness, efficiency and celerity[10]. The data that is used in this study is categorical data from The Female Daily website (femaledaily.com, web forum for skin care) instead of a discreet data, since k-means could handle categorical data that have pre-processed vector to become numerical data. Also, k-means clustered similar typed data so that the recommendation process becomes more relevant because the skin care recommendation need correlation to one another.

K-means clustering used in this research to calculate the products into cluster so that system will give products recommendation from the cluster where the user of the systems has liked the product(s).

3. Results and Discussions

The result of the research is a website which the Entity Relationship diagram (ERD) is presented in figure 1.

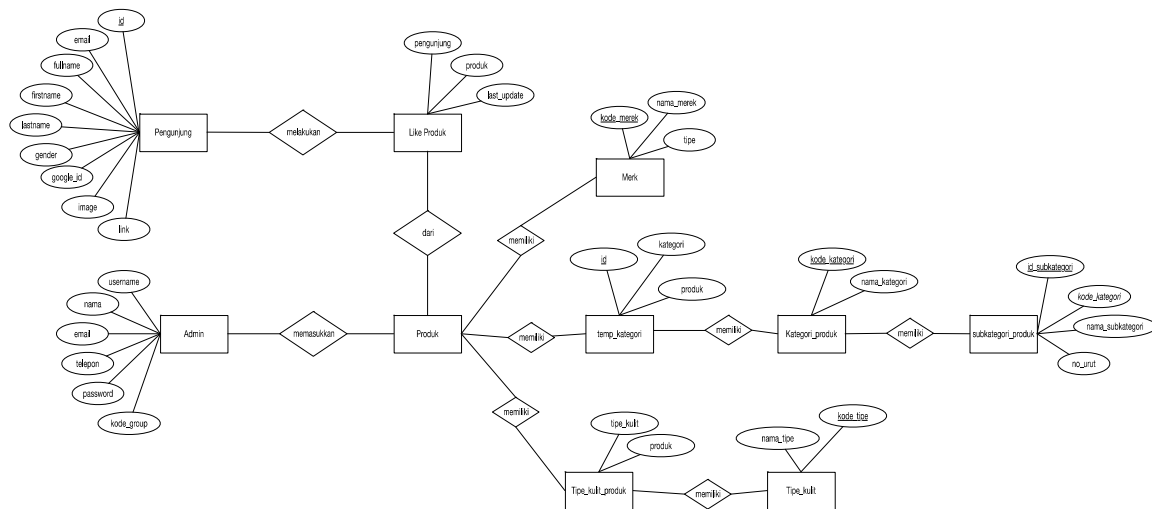


Figure 1: Entity Relationship Diagram

Description of Entity Relational Diagram (ERD):

1. Entitas Pengunjung is a table containing general personal information from system users.
2. Like_Produk entity tables contain product data that users like.
3. Produk is a table that contains information on skin care products.
4. Admin is a table containing information about the administrator of the recommendation system.
5. Temp_category entities
6. Kategori_produk is a table containing information about skin care products.
7. The Subkategori_produk is a table that contains information about skin care product subcriteria.
8. Tipe_kulit_produk entity, namely type_x skin and product table transaction table
9. Tipe_kulit entities are tables that contain the type of product skin type.

The data used in the system is described in the form of Data Flow Diagrams (figure 2).

DFD Level 0

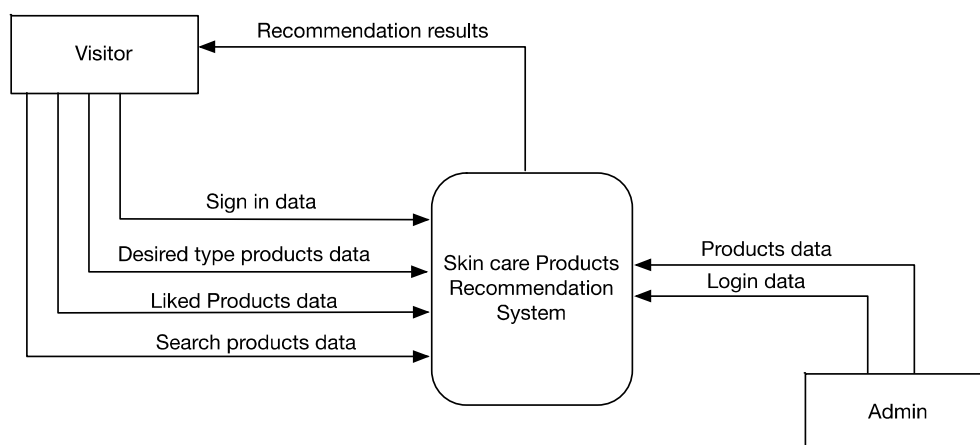


Figure 2: DFD level 0

Figure 2 illustrates all data that flow into and out of the system. There are 2 entities, namely visitors and admin. To be able to access the system, authentication is needed. The sign in process is done with Google account. If the visitor has successfully signed in, they can access the skin care product recommendation system. Next, the admin will input skin care products. Visitors could start liking suitable products and get recommendations from the system.

DFD Level 1

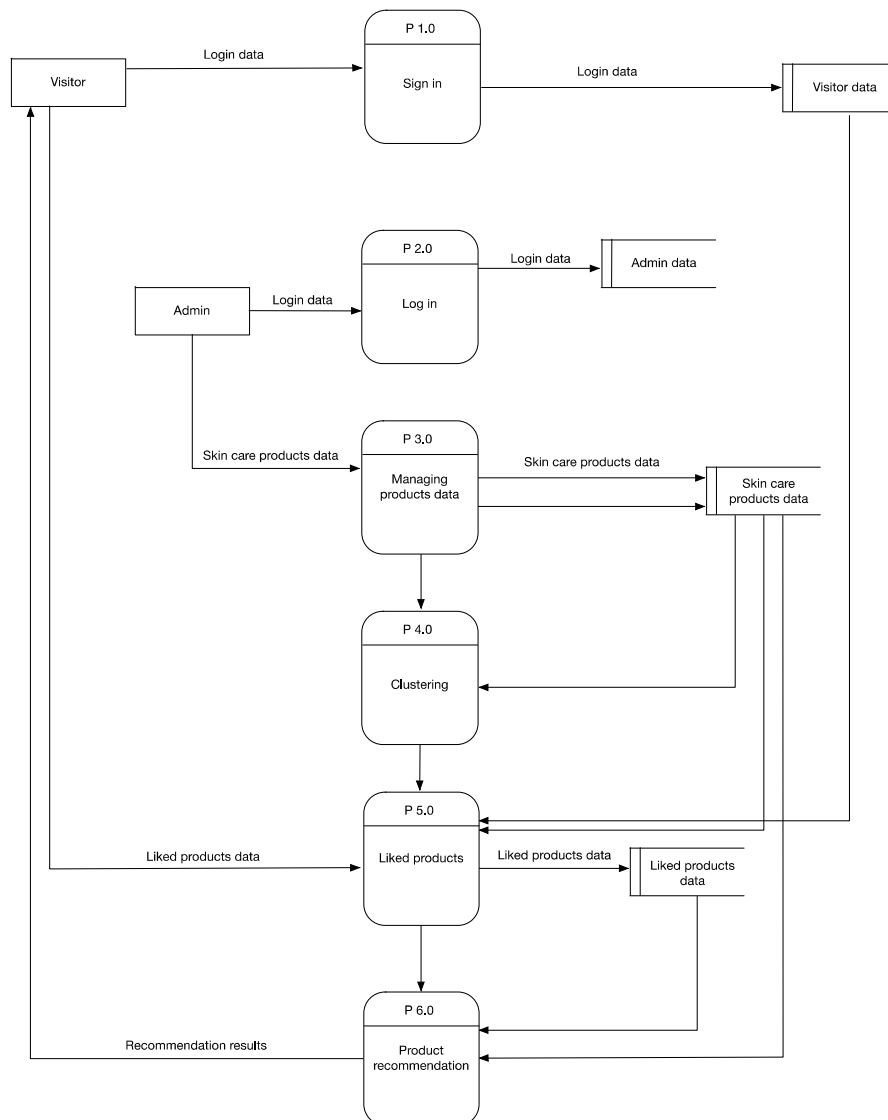
*Figure 3: DFD Level 1*

Figure 3 describes the system in more detail. Description of each process;

- **Process 1**
To access the system and get product recommendations, visitors must sign in first. The sign in process must be done using Google account.
- **Process 2**
The system admin must log in to get access rights to the skin care product recommendation system.
- **Process 3**
If the admin has obtained access rights to the system, the admin can enter data on skin care products that will be recommended through a skin care product recommendation system. Data that must be entered is specified in the system.
- **Process 4**
Product data that has entered the system will be clustered.
- **Process 5**
Users / visitors will like one or more products to get recommendations from the system.
- **Process 6**

The system will perform data clustering calculations for skin care products using Content Based Filtering method. The system will give recommendations to the user of the product cluster that is liked. The system will see if products the user likes are located in the cluster, then the products on the same cluster will be recommended to the user.

The result is web-based system as displayed. There are two actors that can operate this system, Pengunjung (users) and Admin(s)

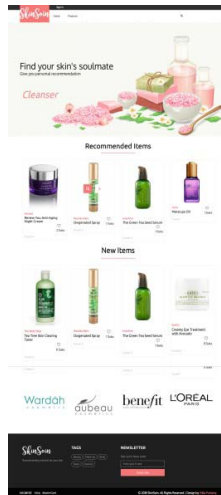


Figure 4: Pengunjung home page (before sign in)

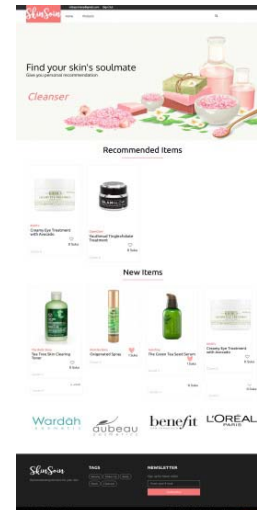


Figure 5: Pengunjung home page (after sign in)

Before sign in, users cannot receive recommendations (figure 4). Then at the recommended products section, the most liked products by other users in the system will be displayed (figure 5). After users sign in with their google account, users need to give like to some products that they have used and suitable for their skin. With this, the recommended products section will show suggested products that might be suitable for the users. While the admin can configure products (figure 6), categories (figure 7), subcategories (figure 8), skin type (figure 9), brands (figure 10), and users (figure 11) from admin menu page.

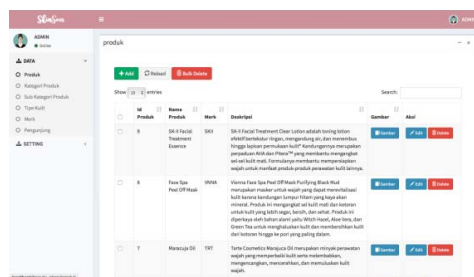


Figure 6: Admin Menu Page: Product

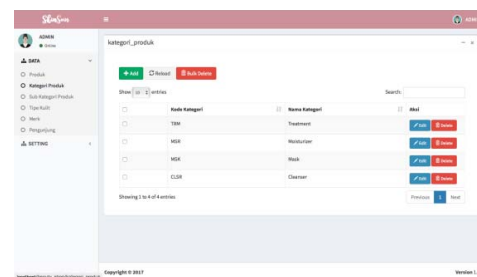


Figure 7: Admin Menu Page: Category

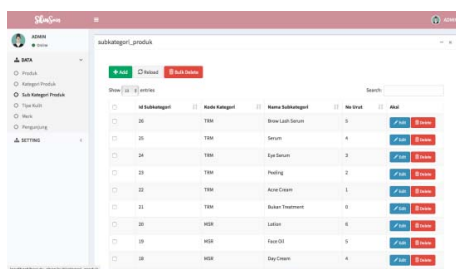


Figure 8: Admin Menu Page: SubCategory

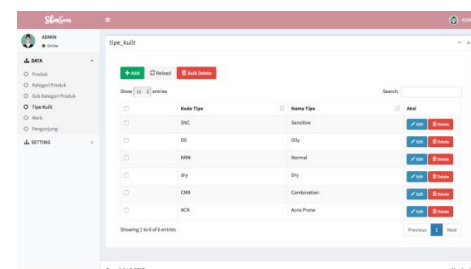


Figure 9: Admin Menu Page: Skin Type

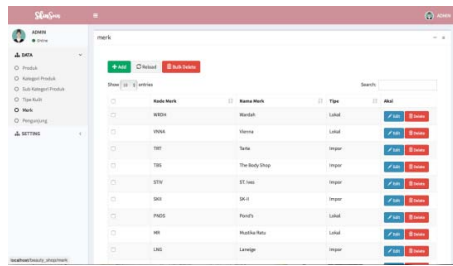


Figure 10: Admin Menu Page: Brand

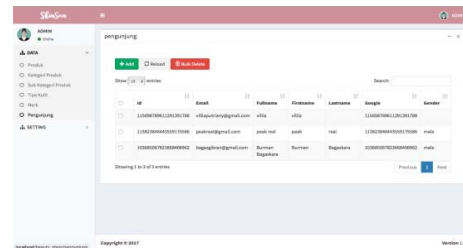


Figure 11: Admin Menu Page: User

4. Evaluation

The cluster was evaluated using Purity to measure the ability of the clustering method. Purity is a method to measure the extent to which clusters contain single class. The calculation thought of as follow:

$$purity(\Omega, \Gamma) = \frac{1}{N} \sum_k \max |\omega_k \cap c_j| \quad (1)$$

Where $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ is the set of clusters and $C = \{c_1, c_2, \dots, c_J\}$ is the set of classes. We interpret ω_k as the set of documents in ω_k and c_j as the set of documents in c_j in Equation (1). In this system, we have five clusters based on skin type of the products for. Purity is $(1/17) * (2+1+0+0+2) = 0.29$.

5. Conclusion

The recommended system can suggest skin care products based on products those had been liked by the user. The product that liked by the user considered as the product that suitable for their skin based on their experience. The recommendation products expected to help user avoid side effects of an unsuitable products.

For the further research in the future, we suggest to develop system that able to solved unexpected problems from users. Also, the system could apply to mobile-based applications so that the users can get recommendations easier.

References

- [1] Anonymous. (2003, May). Pots of promise. The Economist. Retrieved from <http://www.economist.com/node/1795852>
- [2] Łopaciuk, A., & Łoboda, M. (2013). Global Beauty Industry Trends in the 21st Century. Knowledge Management & Innovation Knowledge and Learning, 1079–1087. Retrieved from <http://www.toknowpress.net/ISBN/978-961-6914-02-4/papers/ML13-365.pdf>
- [3] Cheong, S., & Michalski, T. (2016). Asia Personal Care & Cosmetics Market Guide 2016. U.S. Gov. International Trade Administration. Retrieved October 12, 2017, from <https://www.trade.gov/industry/materials/AsiaCosmeticsMarketGuide.pdf>
- [4] Association, A. S. for D. S. (2016). Archive - The 2016 ASDS/ASDSA Annual Report.
- [5] Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). Recommendation Systems Handbook. (F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor, Eds.) (1st ed.). Springer US. <https://doi.org/10.1007/978-0-387-85820-3>
- [6] Son, J., & Kim, S. B. (2017). Content-based filtering for recommendation systems using multiattribute networks. Expert Systems with Applications. <https://doi.org/10.1016/j.eswa.2017.08.008>
- [7] Fahim, Ahmed & M.Salem, Abdel-Badeeh & A. Torkey, F & Ramadan, Mohamed. (2006). Efficient enhanced k-means clustering algorithm. Journal of Zhejiang University SCIENCE A. 7. 1626-1633. 10.1631/jzus.2006.A1626.
- [8] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013.
- [9] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, “A content-based recommender system for computer science publications,” *Knowledge-Based Syst.*, vol. 157, no. February 2017, pp. 1–9, 2018.
- [10] Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>