# Cosmetic Recommendation System based on Ingredients using Cosine Similarity


1st Sekar Alisha Firdaus
*Data Science Program*
*Computer Science Department*
*School of Computer Science,*
*Bina Nusantara University,*
Jakarta, Indonesia 11480
sekar.firdaus@binus.ac.id

2nd Laksmi Dwi Nurmuthia
*Data Science Program*
*Computer Science Department*
*School of Computer Science,*
*Bina Nusantara University,*
Jakarta, Indonesia 11480
laksmi.nurmuthia@binus.ac.id

3rd Raina Imtiyaz
*Data Science Program*
*Computer Science Department*
*School of Computer Science,*
*Bina Nusantara University,*
Jakarta, Indonesia 11480
raina.imtiyaz@binus.ac.id

4th Alexander Agung Santoso Gunawan
*Computer Science Department*
*School of Computer Science,*
*Bina Nusantara University,*
Jakarta, Indonesia 11480
aagung@binus.edu

5th Karli Eka Setiawan
*Computer Science Department*
*School of Computer Science,*
*Bina Nusantara University,*
Jakarta, Indonesia 11480
karli.setiawan@binus.ac.id



***Abstract*—The rapid growth of skincare brands and products can be overwhelming for consumers, making it difficult for them to find products that meet their specific needs. This can lead to frustration, wasted resources, and sometimes even harm from using products with incompatible ingredients. To address this issue, this research introduces a new recommendation system. This system uses a method called 'cosine similarity' to carefully analyze the ingredients in each skincare product. Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. In this context, it is used to compare the ingredient lists of different skincare products. By understanding the complex relationships between different ingredients, the system can provide personalized recommendations that make the selection process easier and help users make informed decisions for better skin health. The effectiveness of this new approach is thoroughly tested through user feedback. Users of the system are asked to provide feedback on the accuracy of the recommendations and whether the recommended products met their needs. This feedback is then used to continually improve the system and ensure it is providing the most accurate and helpful recommendations possible. The research demonstrates its potential to change the way skincare recommendations are personalized in the growing beauty industry. By giving consumers this important knowledge, the research opens up a future where personalized skincare solutions are easily available and accessible to everyone, leading to high levels of customer satisfaction and loyalty for both consumers and brands.**



***Keywords*—Cosine similarity, Recommendation system, skincare, skin types, cosmetics**


## I. Introduction

Cosmetics are a diverse group of products created primarily for external use and intended to wash, scent, change the appearance of, remove odors emerging from, or, more generally, keep the portions of the body to which they are applied in excellent condition. As a result of globalization, the role of these items is fast changing, and their use is widely recognized as a vital component of personal wellbeing [1]. The development in cosmetic products is very good and shows good figures. The global market for cosmetic products was worth USD 532.43 billion in 2017 and is projected to reach USD 805.61 billion by 2023, with a CAGR of 7.14% from 2018 to 2023. The cosmetic industry has grown at a rate of about 15% per year in recent years and continues to do so [2].

Skincare has become a frequent topic of discussion lately, among both women and men. There are numerous skincare brands offering a wide range of products for personal care. When people decide to purchase skincare items, they naturally focus on the product's ingredients and its functions [3]. However, due to the sheer variety of products and brands available, individuals often find it challenging to make a choice. Moreover, making the wrong choice can have adverse effects on the skin, including irritation, breakouts, or allergies. Therefore, there is a need for a way to assist people in finding skincare products that suit their skin types without difficulty [4].

Due to the growing complexity of the skincare market, where options are plentiful, yet appropriate choices remain elusive, it has become a very interesting topic to address by utilizing appropriate recommendation systems and integrating them with existing dermatological principles.

This research aims to solve the complex challenges in skincare through the application of cosine similarity in recommendation systems. This approach combines cosine similarity technology with current dermatology principles to develop a skincare recommendation system based on skin type and skin concerns. This research also delves into the theoretical field of dermatology by formulating and testing new theories based on in-depth data analysis. Testing on user persona is a part of this research, ensuring that the resulting recommendation system can provide relevant product to individuals according to their skin needs.

This research can certainly help many people. Consumers, for instance, will benefit from it. By providing recommendations for ordinary consumers, it helps them choose effective skincare treatments that are suitable for their skin types. In this regard, the beauty industry also reaps its own benefits. They can develop products that are more suitable for different skin types and increase their sales by promoting their products to the right customers. Other beneficiaries include software developers, who can create recommendation applications and technologies and offer them to the beauty industry to boost their sales.

The data is taken from Kaggle. After the data collection process, there is a data preprocessing process, and then built the model. The data analysis that we used is classification. The cosine similarity model is used to provide recommendations based on the similarity of ingredients. Then, the model is tested using a user persona in several cases. This paper follows a conventional structure, comprising several chapters. In Chapter I, Introduction, it contains general knowledge of the topic we raised, our purpose of doing this research, and the benefits of this research to the wider community. In Chapter II, Related Works, contains previous research that has been done related to the topic. Chapter III, Research Methodology contains an explanation of the methods that has been used. This chapter contains detailed information about data collection and steps to create a recommendation system. Chapter IV, Results and Discussion contains an explanation of the results of the research. The contents of this chapter are to answer the objectives that are described in Chapter I. Chapter V, Conclusion, summarizes the research conducted and provides a clear conclusion. The paper concludes with a section for references, which catalogs all the sources, literature, and references cited within the paper, adhering to IEEE citation style.

## II. Related works

Young ladies frequently have various skin-related concerns. Many of them still do not know their skin types or which components to use to cure and progressively improve their skin issues. Because our skin serves as the biggest defense against illness, proper skincare is crucial. Healthy skin helps maintain the integrity of this barrier. That is what will be explored in this paper. To understand which skincare products are suitable for different skin types [5].

Firstly, Swati Solanki and Gayatr Jain (Pandi) proposed a predictive system that provides a precise idea of which product is best for all skin types using the Deep Learning Technique. Their work focuses on developing a system that can effectively improve the recommended effectiveness of cosmetic products for different skin types. The technique they used is DNN or Deep Neural Network. They implemented a DNN model with two hidden layers, with 70 nodes in the dense layer, activation function ReLU, and 0.15 dropout ratio. Then, they added an output layer with SOFTMAX function with three nodes for multiclass classification. They have trained their model with 100 epochs and batch size of 10 and tested the results of the model using accuracy metrics. In conclusion, the proposed system suggests the products with 85.60% accuracy [6].

In another study, Muskan Chaurasia and his colleagues conducted their own research on a system that suggests skincare and makeup products that suitable for a customer's skin type. They used t-distributed Stochastic Neighbor Embedding (t-SNE) in this research. The model states that if a customer does not know which product they should buy, they can insert their desired beauty impact instead of the name of the product's name. The t-SNE algorithm is used to cluster similar products together based on their ingredients and other characteristics. The proposed system made by the algorithm helps customers to find the right product based on their skin type and beauty desires [7].

Then, there are Hsiao-hui Li and his colleagues who performed image processing using the YOLOv4 algorithm to detect key features in facial images and intercept sub-images of the Region of Interest (ROI) as input information for multi-label models. Their focus is on applying machine learning and deep learning algorithm development to a human face and skin intelligence recommendation platform to provide product recommendations that match the user's skin condition. The first step is image recognition, labeling, and feature extraction. Next, the YOLOv4 algorithm is applied to generate a recommendation platform from the prediction results [8].

Mina and her colleagues applied the Simple Additive Weighting (SAW) method to provide skincare recommendations for various skin types. The SAW method was used as a basis for compiling basic skincare product recommendations, taking into account the characteristics of each skin type. This research made an important contribution in helping consumers choose skincare products that suit their skin type through the SAW algorithm approach [9].

Furthermore, Diana and her colleagues used the methods of image recommendation, speech recognition, text mining, and time series data prediction to create a recommendation system for predicting items based on user opinion. This system helps customers make informed decisions about the products they buy. This research is expected to explore deep learning algorithms to improve the accuracy of recommendation system [10].

There is also Samrat Ray et al. that used Deep Learning to make a cosmetics suggestion system. They used convolutional neural network for the system. The proposed deep learning approach transfers input attributes such as skin type and component compatibility to input layers before propagating to hidden layers. Then, the cosmetic combination is shown on the output layer. Alternative configuration strategies outperform conventional ones. Their model is successful in recommending cosmetics for various skin types, with an accuracy rate of 97.38% [11].

In a similar vein, Arya Kothari et al. Also used convolutional neural network to make classification for cosmetic skin type. Their dataset consisted of over 80 skin images collected by web scrapping. They used MTCNN model to detect and extract facial features. Then the CNN model built to predict the skin type. After training for 60 epochs with learning rate of 0.0003, there is a change in the accuracy and the loss. The trend depicts how the model's accuracy and loss decreased steadily with each epoch. The final accuracy found was approximately 92%, with a loss of 0.31 [12] .

Lastly, Shwetambari Borade et al. used machine learning to detect and diagnosis the type of cosmetic skin and its disease. They made this automated methods-based applications to make a diagnosis in the early stage. The input is of an image then it will be processed. The RGB image values are converted to HSV and YCbCr value systems, and the HSV as well as YCbCr values of each pixel are attributed to the standard values of a skin pixel. After processing the image, they did segmentation for identifying the skin diseases. Last, they did the classification using random forest classifier. The resuls show good value, with an accuracy rate of 95% [13].

## III. Research methodology

### A. Data Collection and Preprocessing

In this paper, the data is taken from Kaggle [14]. The dataset is a cosmetic dataset. This dataset contains information about whether a product is suitable for a particular skin type. The skin types in question are oily, dry, normal, and combination. The data from this dataset is collected from Sephora and the total of products is 1472 products. This dataset has eleven columns. These columns consist of the type of product, brand of product, name of cosmetic, price, ranking, ingredients, and whether the product suits dry, normal, oily, or combination skin.

To gain further insights into the distribution of products across different skincare labels, a visual representation of the data is created.

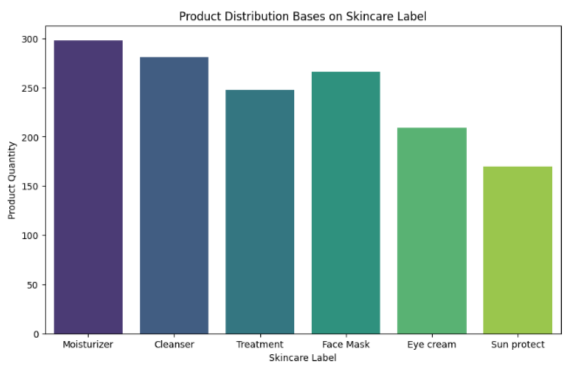


Fig. 1. Product Distribution of Skincare Category

As shown in the figure, the number of products varies considerably depending on the label. Moisturizers take the lead with nearly 300 products in the dataset. Cleansers follow closely behind with over 250 products. Followed by facemask which also over 250 products, then treatment which almost reached 250 products, eye cream which over 200 products, and lastly, sun protection which over 150 products.

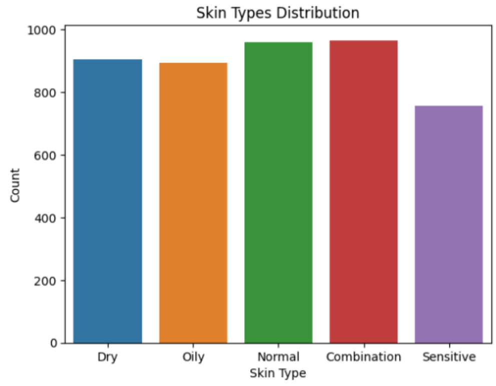


Fig. 2. Product Distribution of Skin types

Based on the figures above, the results are similar. For dry skin types, the number is between 800-1000 products. For the oily skin type, it is very similar to dry skin, but the value is below dry skin. For normal skin type, the value is close to 1000. For combination skin type, it is also close to 1000 but slightly more than normal skin type. For sensitive skin type, it has the lowest value, below 800.

## B. *Label Segmentation*

The data is segmented into specific skincare labels such as moisturizer, cleanser, face mask, treatment, eye cream, and sun protect. Each label is further categorized according to different skin types, such as dry skin, normal skin, combination skin, oily skin, and sensitive skin. This segmentation allows detailed analysis of ingredients based on specific product categories and skin types.

## C. *Document Term Matrix (DTM) Creation*

To facilitate the calculation of cosine similarity, a document term matrix is created for each combination of skincare label and skin type. Ingredient information's are encoded into binary representation. The ingredients listed in each cosmetic product were processed to create a standardized representation for analysis. The text in the ingredient column has been converted to lowercase and tokenized. A document term matrix was created for each combination of skincare label and skin type to represent the presence or absence of certain ingredients [15]. Each ingredient is assigned a unique index for efficient representation within the matrix. The resulting DTM serves as input to the cosine similarity calculation.

## D. *Cosine Similarity Calculation*

Cosine similarity is a way to measure the similarity between one vector and another vector. Cosine similarity measures the angle between these two vectors. More clearly, the dot product between the two vectors is divided by the product of the magnitude. The value of cosine similarity can be two values, namely 0 and 1. If the value is 1, it means that these two vectors are similar, while if the value is 0, these two vectors are not similar at all [16].

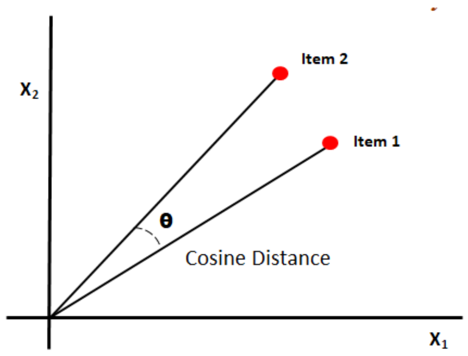


Fig. 3. Visual illustration of Cosine Similarity

$$\cos(\theta) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} Ai\,Bi}{\sqrt{\sum_{i=1}^{n} Ai^2}\sqrt{\sum_{i=1}^{n} Bi^2}} \quad (1)$$

A = Value of vector A
B = Value of vector B

To calculate the cosine similarity score in this research, the A and B values are the ingredients vector values of one skincare product based on the skincare category and skin type. One object in the cosine similarity matrix refers to one skincare product. If the cosine similarity score is 1, it means that the product A and product B that are compared have the same ingredients.

In this research, the cosine similarity matrix is created for each product in the skincare category grouped by skin type. In the data, there are 6 skincare categories and 5 skin types, so the total matrix created is 30 with various matrix sizes according to the number of products in each skincare category based on skin type.

The recommendation system relied on cosine similarity, a similarity measurement between two non-zero vectors. Cosine similarity was calculated based on the ingredients of different cosmetic products, allowing to identify products with similar ingredient compositions. This approach ensured a personalized recommendation system that considers both skincare label and skin type. The average cosine similarity score is then also calculated to assess the overall similarity within each skincare category. This involved the creation of similarity matrices for individual products considering both skincare label and skin type. The upper tringular part of each matrix (excluding the diagonal) is used to calculate the average similarity score, providing insight into the overall similarity of ingredients within each skincare category.

## E. *Product Recommendation*

A product recommendation algorithm has been developed that suggests cosmetics based on the purchased item/s. The algorithm considered the skincare label, skin type, and ingredient. Given a specific product type, skin type, and purchased product as an input, the algorithm calculated cosine similarity to other products and recommended the most similar product, excluding the purchased product itself.

**Algorithm 1: Pseudocode for building skincare recommendation system using cosine similarity**

***Inputs:*** *Cosmetics dataset*
***Output:*** *Skincare recommendation*
*1. **Start:***
*2. Import Libraries*
*3. Load the Cosmetics dataset*
*4. Perform Exploratory Data Analysis*
*5. Categorize skincare category based on skin type*
*6. Tokenize the ingredients for each skincare category and skin type*
*7. Create document-term matrix (DTM) for each skincare category and skin type*
*8. Create dictionary of dataframes and DTM for each skincare category and skin type*
*9. Calculate the cosine similarity for each skincare category and skin type*
*10. Calculate the average cosine similarity for each skincare category*
*11. Define a function to recommend a skincare based on the user's input:*
*12. Input: Skincare category, skin type, and purchased skincare*
*13. Find the dataframe and DTM for the specific skincare category and skin type*
*14. Get the index of the purchased skincare in the DTM*
*15. Calculate the cosine similarity matrix for the skincare*
*16. Get similarity score for the purchased skincare*
*17. Sort products based on similarity from most similar to least*
*18. Return the name of the most similar product (excluding the purchased product itself)*
*19. Call the recommend_product function with the required inputs and print the recommended product*
*20. **End***

## IV. Result and Discussion

The research conducted has applied cosine similarity in a recommendation system. The objective was to assist consumers in choosing effective skincare based on their skin type. This research incorporated insights from established dermatological principles and combined them with advanced technology to formulate a personalized skincare recommendation system.

The methodology of this research included data collection from Kaggle, preprocessing, and development. The data, which consisted of information on various skincare products and their suitability for different skin types was analyzed using cosine similarity. The cosine similarity model was used to provide recommendations based on ingredient similarity, ensuring a personalized approach that considers skincare labels and skin types.

TABLE I. COSINE SIMILARITY EXAMPLE ON MOISTURIZER WITH DRY SKIN TYPE

| | 0 | 1 | 2 | ... | 189 |
|---|---|---|---|---|---|
| 0 | 1.000000 | 0.116642 | 0.060783 | ... | 0.0 |
| 1 | 0.116642 | 1.000000 | 0.148888 | ... | 0.0 |
| 2 | 0.60783 | 0.148888 | 1.000000 | ... | 0.0 |
| ... | ... | ... | ... | ... | ... |
| 189 | 0.0 | 0.0 | 0.0 | ... | 1.000000 |

The table above is one of several cosine similarity matrices created, which displayed the cosine similarity matrix of products in the moisturizer skincare category for dry skin, with the total of 189 products. The matrix shows that the same products (object 0 and 0, 1 and 1, 2 and 2, and so on) have a perfect cosine similarity score of 1.000000, while for different products show various cosine similarity scores which indicate the difference in ingredients of the two products being compared.

The results obtained from the cosine similarity analysis have shown the degree of similarity among the products in each skincare category. By examining the individual matrices, it has been observed that certain products have exhibited higher cosine similarity scores, indicating a greater similarity in their ingredient compositions and thus indicating potential clustering of products within each category, with some products showing closer alignment in terms of ingredients. To provide a more comprehensive picture of the overall similarity within each skincare category, an average cosine similarity score was calculated.

The average cosine similarity scores calculated for each skincare category are as follows:

TABLE II. AVERAGE COSINE SIMILARITY SCORE

| Category | Average Cosine Similarity Score |
|---|---|
| Moisturizer | 0.139 |
| Cleanser | 0.117 |
| Face Mask | 0.153 |
| Treatment | 0.145 |
| Eye Cream | 0.158 |
| Sun Protect | 0.122 |

The computed average cosine similarity scores present a quantifiable measure of the average similarity between products within each skincare category. The eye cream category in particular stands out with the highest average cosine similarity score of 0.158, indicating a relatively higher level of similarity among the ingredients of products in this particular category. In contrast, the cleanser category showed the lowest average cosine similarity score of 0.117, which signifies a lower average similarity among the ingredients of the various cleansing products.

In table 2, it can be seen that the average cosine similarity score of each skincare category is quite small even though it is only calculated based on the same skincare category. The average cosine similarity score will get smaller if there are many small scores in the cosine similarity score. The cosine similarity score is small because it is rare to find similarities in the ingredients of several skincare products. Because of that, this research focused on recommending skincare products that have the same skincare category as the products already purchased. This helps users if the product that is often purchased is out of stock or even discontinued, users can buy products with the same skincare category with similar ingredients.

The implementation of a recommendation system based on cosine similarity scores aimed to assist consumers in identifying products that are similar to those they have previously purchased. However, it is important to recognize instances where the system returns a ‘Product Not Found’ message. This indicates that the combination of product category, skin type, and specific product purchased did not have enough data to generate appropriate recommendation.

To provide skincare recommendation results, this recommendation system requires input which consists of skincare category, skin type, and purchased product where the purchased product is assumed to work well on the user's skin. After entering the input, the system will issue an output which is a product recommendation with the same skincare category as the purchased product.

TABLE III. SKINCARE RECOMMENDATION RESULTS

| No | Skincare Category | Skintype | Purchased Product | Recommendation |
|---|---|---|---|---|
| 1. | Eye cream | Oily | Eye Treatment | Product Not Found |
| 2. | Cleanser | Dry | Essential Power Skin Toner for Normal to Dry Skin | Time Release Acne Cleanser |
| 3 | Treatment | Combination | The Treatment Lotion | The Revitalizing Hydrating Serum |
| 4 | Moisturizer | Combination | The Water Cream | CC Cream Daily Correct Broad Spectrum SPF 35+ Sunscreen |
| 5 | Face Mask | Sensitive | Hydrating Facial Sheet Mask | WHITE Brightening Mask |

The table shows the application of this recommendation system. When exploring the category of eye cream for oily skin where the purchased product is Eye Treatment, the system returned the message 'Product Not Found'. For the cleanser category for dry skin with purchased product 'Essential Power Skin Toner for Nornal to Dry Skin' , the system successfully recommended Time Release Acne Cleanser. Then, for the treatment category for combination skin with purchased product 'The Treatment Lotion', the system recommends The Revitalizing Hydrating Serum. For the moisturizer category with combination skin and purchased product 'The Water Crean', the system recommends CC Cream Daily Correct Broad Spectrum SPF 35+ Sunscreen. Lastly, for face mask category with sensitive skin and purchased product 'Hydrating Facial Sheet Mask', the system recommends WHITE Brightening Mask. These examples highlight the system's ability to provide appropriate recommendations while underscoring the importance of data availability for more accurate and relevant suggestions.

The contribution of this research is clearly seen in the explanation above, where average cosine similarity scores were calculated for different skincare categories. The findings revealed various degrees of ingredient similarity within each category, with eye cream standing out with the highest average cosine similarity scores. The implementation of the recommendation system successfully assisted consumers in identifying products that were similar to those they had previously purchased.

However, the availability of data to generate accurate recommendations needs to be emphasized. The instances where the system returned a 'Product Not Found' message underscored the need for sufficient data to generate appropriate and relevant recommendations. The examples given highlight the ability of the system to provide appropriate recommendations while emphasizing the importance of data quality and completeness.

## V. Conclusion

In conclusion, this research has demonstrated the potential of cosine similarity in developing a skincare recommendation system that considers skincare label and skin type. The findings provide valuable insights into inter-product relationships within the skincare segment and offer a quantifiable measure of ingredient similarity. This research not only benefits consumers in choosing effective skincare products, but also has implications for the beauty industry and software developers looking to improve their product offerings. There are several related works, which showcase various approaches to skincare recommendation systems. Researchers explored techniques such as deep learning, t-SNE algorithm, YOLOv4 algorithm, Simple Additive Weighting (SAW) method, and a combination of image recommendation, speech recognition, text mining, and time series data prediction to improve skincare product recommendations.

Future research is expected to develop the existing model, such as creating a recommendation system with different categories from purchased product. In addition, future research could consider incorporating other factors such as price, discounts, and so on into the model. By considering these factors, the recommendation system will become more relevant and can provide recommendations that are more in line with the needs and financial capabilities of users. This will ultimately increase user satisfaction and product sales.

## References


[1] L. Salvioni et al., "The emerging role of nanotechnology in skincare," Advances in Colloid and Interface Science, vol. 293, p. 102437, Jul. 2021.

[2] M. Bilal and H. M. N. Iqbal, "New Insights on Unique Features and Role of Nanostructured Materials in Cosmetics," Cosmetics, vol. 7, no. 2, p. 24, Apr. 2020.

[3] K.B. Nalina, A. Adarsh, and A. Puttabuddhi, "Consumer Awareness for Ayurvedic SkinCare Products," *International Research Journal on Advanced Science Hub*, vol. 5, no. 08, pp. 257-268, Aug. 2023.

[4] Campen, Rebecca B. The Comprehensive Guide to Skin Care: From Acne to Wrinkles, What to Do (and Not Do) to Stay Healthy and Look Your Best. ABC-CLIO, 12 Nov. 2009.

[5] N. Ramlan, M. Hafiz, F. Kamil, D. Ghani, A. Amira, and A. Rahman, "Journal of critical reviews 900 journal of critical reviews MOBILE APPLICATION DESIGN PROCESS TO ACHIEVE HEALTHY SKIN: A GUIDELINES AND FRAMEWORK," vol. 7, p. 2020.

[6] S. Solanki and G. Jain, "Deep Learning Technique for selecting appropriate Beauty Care Products for different skin Type," *Journal of emerging technologies and innovative research*, Jun. 2020.

[7] M. Chaurasia et al., "A Machine Learning Based Recommendation System For Cosmetics," *Journal of Pharmaceutical Negative Results*, pp. 3711-3716, 2022.

[8] H.H. Li et al., "Based on Machine Learning for Personalized Skin Care Products Recommendation Engine," *in 2020 International Symposium on Computer, Consumer and Control (IS3C)*, 2020.

[9] M.I. Rahayu et al., "SISTEM REKOMENDASI PEMILIHAN PRODUK BASIC SKINCARE BERDASARKAN JENIS KULIT MENGGUNAKAN ALGORITMA SIMPLE

ADDITIVE WEIGHTING (SAW)," *Jurnal Teknologi Informasi Dan Komunikasi*, vol. 11, no. 2, pp. 1–6, 2022.

[10] D. Ikasari et al., "Metode Deep Learning Pada Sistem Rekomendasi : Review Paper," *Prosiding Seminar SeNTIK*, vol. 2, no. 1, pp. 42–49, 2018.

[11] S. Ray, A. M, A. K. Rao, S. K. Shukla, S. Gupta and P. Rawat, "Cosmetics Suggestion System using Deep Learning," *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, 2022, pp. 680-684, doi: 10.1109/ICTACS56270.2022.9987850.

[12] A. Kothari, D. Shah, T. Soni and S. Dhage, "Cosmetic Skin Type Classification Using CNN With Product Recommendation," *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2021, pp. 1-6.

[13] S. Borade, D. Kalbande, H. Jakaria and L. Patil, "An Automated approach to detect & diagnosis the type of Cosmetic Skin & its Disease using Machine Learning," *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT*), Bangalore, India, 2022, pp. 1-8.

[14] A. A. Awan, "Cosmetics Datasets," Kaggle, 2021.

[15] Anak Agung Putri Ratna, F. Astha Ekadiyanto, Mardiyah, Prima Dewi Purnamasari, and M. Salman, "*Analysis on the Effect of Term-Document's Matrix to the Accuracy of Latent-Semantic-Analysis-Based Cross-Language Plagiarism Detection*," Dec. 2016.

[16] S. Siddika and M. S. Hossen, "Automatic Text Summarization Using Term Frequency, Luhn's Heuristic, and Cosine Similarity Approaches," *in 2022 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET)*, 2022, pp. 1-6.