Original Article

A Machine Learning Based Recommendation System For Cosmetics

Muskan Chaurasia[1] Neha Pathak[2] Meetu Rani[3] Muskan Verma[4] Nandini Gauhri[5]

[1]UG Scholar, Computer Science & Engineering, Meerut Institute of Engineering and Technology, Meerut, India. Email: muskanchaurasia.2024@gmail.com
[2]UG Scholar, Computer Science & Engineering, Meerut Institute of Engineering and Technology, Meerut, India. Email: imnehapathak0504@gmail.com
[3]Assistant Professor, Computer Science & Engineering, Meerut Institute of Engineering and Technology, Meerut, India. Email: meetumann@gmail.com
[4]UG Scholar, Computer Science & Engineering, Meerut Institute of Engineering and Technology, Meerut, India. Email: verma24.muskan@gmail.com
[5]UG Scholar, Computer Science & Engineering, Meerut Institute of Engineering and Technology, Meerut, India. Email: nandinigauhri1@gmail.com
DOI: 10.47750/pnr.2022.13.S10.449

Abstract

The demand for cosmetics has grown recently, especially in the area of skincare, around the globe. Consumers have traditionally relied on top-selling items or suggestions from the counter while shopping in-store. These are not reliable ways to determine if a product will work with a particular user because everyone has a distinct skin condition. The main goal of this proposal is to develop a system for recommending skincare products based on the user's skin type and the makeup of the product. To determine a product's chemical makeup and locate products with comparable constituent compositions, content-based filtering is utilised. If a user doesn't know what a product is or hasn't discover they like, they may also enter their desired beauty impact instead of a product name using this approach.

Keywords— Recommender system, content-based filtering, ingredients and cosmetics.

I. INTRODUCTION

As a title state itself the method is used to recommend the cosmetic product on the basis of ingredients present in it. As more people began visiting the cosmetics counter to receive product suggestions, there was a corresponding increase in the requirement for advanced technologies. However, this method is time consuming and frequently unsuccessful. Users have found it challenging to make the right decisions as a result of the overwhelming amount of information that is available online. It is seen to be beneficial that there is a wealth of product information and reviews. However, it also limits users from selecting pertinent data and making decisions in accordance with their needs. The urgent demand for tailored solutions that might make it easier to obtain data has been sparked by this problem.

To address the issue of information overload and streamline the decision process, researchers have suggested several recommender systems [7]. Collaboration-based filtering and content-based filtering are the two approaches that are used the most frequently. Recently, a hybrid strategy that combines the two strategies was developed in an effort to optimise the advantages of both approaches while addressing their shortcomings.

Which method is most effective for determining how well a product is suited to a particular buyer is still up for debate. In spite of the fact that each customer's skin condition is different, many online cosmetics retailers nevertheless suggest bestsellers to clients. The recommender systems for personal care goods therefore require additional research and development.

This suggestion offers an approach to recommendations that is content-based and assesses how comparable products' constituent compositions are. The new algorithm makes product suggestions across numerous categories as opposed to only just one category to provide more useful recommendations. Additionally, it offers the user the choice to provide as little information as possible to receive skincare product recommendations. By comparing the system's performance to the product ratings, it will be proven to be accurate.

It is a complicated matter to make a cosmetic recommendation for skin problems. There may be a small group of persons with remarkably identical cosmetic preferences. Additionally, by using user collaborative filtering, we are able to make product recommendations based on the ranking metrics of this nearby group. A more delicate and difficult issue than just suggesting your evening's movie is a person's skin type and facial features. Focusing on the actual contents of each product, or its ingredients, and drawing comparisons based on them is necessary to ensure the recommendation's dependability and stability.

II. LITERATURE REVIEW

This paragraph will contain the prior research that has been done earlier. Here, we will discuss upon the research that has been done on the cosmetic recommendation based on similarities in products, ingredients, reviews and ratings, skin type filtering. Collective filters make use of user-provided data like purchase, likes, clicks, etc. Despite having a cold-start issue, they perform well when given enough data information [1]. Collaborative filtering studies think that finding similarity measure can assist match customers with the right product ideas.

Both Matsunami and Okuda used the method of determining user similarity to examine reviews of cosmetic products [9, 10]. They extracted not only ranking but also actual reviews with individual preferences and opinions using algorithmic grading and k-means cluster analysis [8,12]. Ye also employed collaborative filtering, but she concentrated on addressing the shortcomings of the conventional approach [2,13]. Ratings are used to confirm the results even though it doesn't filter items based on them.

In order to adapt skincare product recommendations quite as much as possible, Putriany used content-based mapping in their study [3,14]. The system was designed around the certain user. She concentrated on the user account but also took into account aspects like skin tone, utilization, pricing, description, and photographs for better suggestion [4,15].

In order to solve this issue, Jeong created a recommender system that is based on the components of cosmetics [5,16,17].

She reasoned that because people's skin kinds and components are more intricate and delicate, suggesting skincare items should be kept apart from promoting movies. Like Jeong, Honma et al. adopted a strategy to link different skin types to different cosmetic chemicals [6,7,11].

III. GENERAL ARCHITECTURE

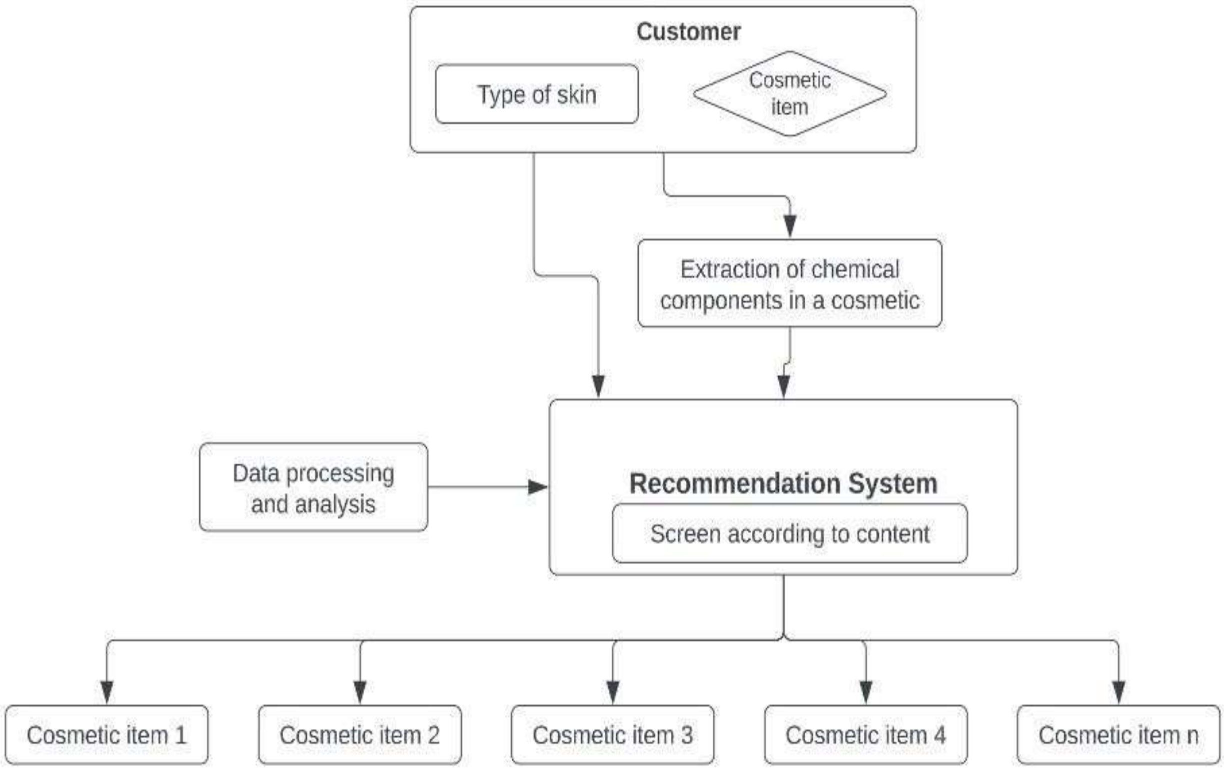

Figure 1: General Architecture

Depending on the item the user enters at the initial stage, the suggested system provides content-based filtering. A user chooses an item from among six categories (moisturiser, cleanser, facial treatments, face mask, sun protection and eye treatment) and specifies one of five skin tones (combination, dry, oily, normal and sensitive) [18,19,20].

The information includes cosmetics from many companies. Only six of the many individual care types were chosen in order to concentrate on beauty products. These 6 categories including cleansers, facial mask, eye treatment, moisturizers, sun screen cream, and facial treatments. The dataset comprises of 1472 products and contains details about each item's name, brand, rank, price, skin type, and chemical ingredients of product [21,22,23,24].

While components are taken from the item, the skin type is directly mapped to the recommendation system. Then, data, which contains details on other goods, is given to the content-based recommendation systems along with the user's skin tones and the item's ingredients [25,26].

With this approach, product recommendations are given for each of the six categories. The system will generate k numbers of recommendations to each of the n product kinds after analysing the comparison of constituent composition among items.

IV. METHODOLOGY

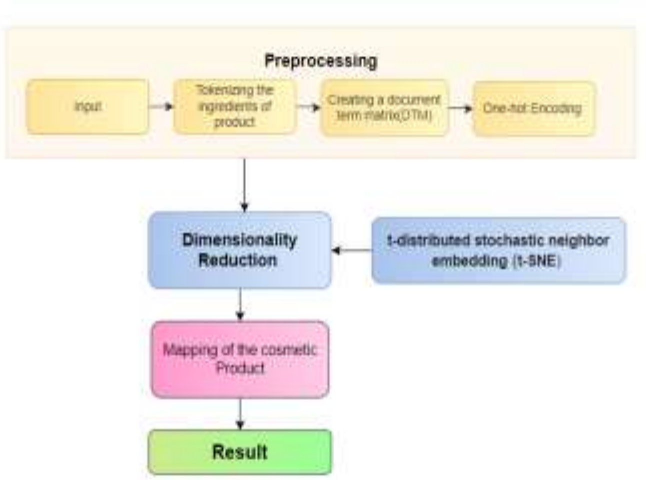

Figure 2: Flow Chart of methodology

Our data includes five various kinds of skin (dry, normal, combination, sensitive and oily) and six product categories (eye creams, face masks, cleansers, moisturizers, and suns cream). Considering that each person's skin type and product needs are unique.

We must first perform certain pre-processing activities and bookkeeping of the original words in each product's components list before we can reach our final objective of comparing the contents in each product. Tokenizing the list of components in the component column will be the first stage. We'll create a binary collection of words after tokenizing them. A dictionary will then be made using the components.

Each cosmetic item will here be compared to a doc, and each chemical component will be compared to a term. As a result, the matrix might be considered a matrix of "cosmetic ingredients." We will first generate an empty matrix initializing with 0. The overall number of cosmetic items in the data is represented by the rows in the matrix. The overall number of ingredients is represented by the columns in the matrix.

We construct a method to calculate total number of tokens for each row. The matrix must be filled with either 1 or 0; if an ingredient appears in a cosmetic product, the value will be 1, otherwise it will remain 0. We will now update the values for each row of this matrix by using the method to the tokens in the corpus. Thus, the outcome will reveal the components of each product. For instance, the results of a cosmetic product including glycerine, water, propanediol, butylene glycol, betaine and so on will be shown here.

	Water	**Propanediol**	**Butylene Glycol**	**Betaine**	**…**	**Hexyl-decanol**
P1	1	1	0	0	…	0
P2	1	0	1	1	…	1
…						

Figure 3: Product-Ingredients Matrix

The updated matrix has the dimensions (190, 2233), which indicates that our data has 2233 features.

We must reduce it into 2D for visualization. Here, we'll apply t-SNE to reduce the data's dimensions. The nonlinear dimensionality reduction algorithm t-distributed Stochastic Neighbour Embedding (t-SNE) is useful for embedding high-dimensional data for visualization in a low-dimensional environment of 2D or 3D dimensions. By reducing the dimension of data while maintaining the commonalities between the cases, this technique can decrease the dimension of data specifically.

As a result, we can vectorize and create a plot on the two - dimensional plane. The distances between the points will show the commonalities between the beauty products in our data, which will vectorized into 2D coordinates.

We may plot all of our things on the two - dimensional plane using the t-SNE values. The best part about this is that it will display the brand name, product name, price, and ranking of each item. Let's use Bokeh to create a scatter graph and add a hovering element to display that data.

V. EXPERIMENTAL RESULT

Analyse the result using the graph we have created. The many plot points are represented by the various cosmetic products. A t-SNE plot's axes are difficult to understand in the context of the initial data. The t-SNE visualization technique allows you to plot high to low dimensional data environment. As a result, a quantitative interpretation of an t-SNE plot is undesirable.

Instead, we may determine the distance between the points on the graph (which points are near and which are far). The two objects' compositions are increasingly similar the closer apart they are. As a result, this allows us to examine the products without having any prior knowledge of chemistry.

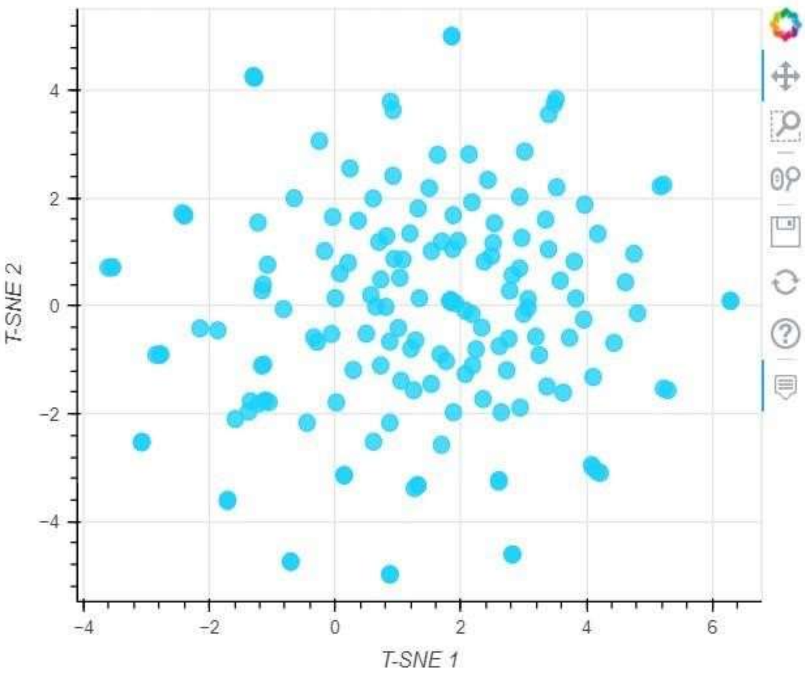

Figure 4: The above graph showing the similarity between the products

VI. CONCLUSION

Any kind of scope can use recommendation focuses on product features. It might also apply to a suggestion for a wine or a book. With the characteristics of the product, we may create numerous charts. We can more easily comprehend the relationships between the things by visualizing them on a map.

We can go closer to providing services that are more precise using this analysis. It could display the product's pictures and customer reviews or offer a straight link to the purchase page. An alternative layout with axes that show chemical properties is also an option. We can adjust the axes for water retention or toxic effect, for instance, if we add any expert chemical information. Then, it will then be able to determine the level of toxicity of each product. These applications will provide a more in-depth level of product evaluation.

VII. REFERENCES

[1] Hansson, Linda. "Product Recommendations in E-commerce Systems using Content-based and Collaborative Filtering." (2015). download (lu.se)

[2] Ye, Hongwu. "A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map." J. Softw. 6.4 (2011): 732-739.

[3] Putriany, Villia, Jaidan Jauhari, and Rahmat Izwan Heroza. "Item Clustering as An Input for Skin Care Product Recommended System using Content Based Filtering." Journal of Physics: Conference Series. Vol. 1196. No. 1. IOP Publishing, 2019.

[4] Joanna Cristy Patty, Elika Thea Kirana, and Made Sandra Diamond Khrismayanti Giri. 2018. Recommendations System for Purchase of Cosmetics Using Content Based Filtering. International Journal of Computer Engineering and Information Technology 10, 1 (Jan. 2018), 1–5. www.ijceit.org

[5] Jeong, Jiwon. "For Your Skin Beauty: Mapping Cosmetic Items with Bokeh." (2018).

[6] Nakajima, Y., Honma, H., Aoshima, H., Akiba, T., & Masuyama, S. (2019, October). Recommender system based on user evaluations and cosmetic ingredients. In 2019 4th International Conference on Information Technology (InCIT) (pp. 22-27). IEEE.

[7] Ndengabaganizi Tonny James and K. Rajkumar. 2017. Product Recommendation Systems based on Hybrid Approach Technology. International Research Journal of Engineering and Technology 4, 8 (Aug. 2017). www.irjet.net

[8] Asami Okuda, Yuki Matsunami, Mayumi Ueda, and Shinsuke Nakajima. 2017. Finding Similar Users Based on Their Preferences against Cosmetic Item Clusters. (2017). https://doi.org/10.1145/3151759.3151829

[9] Narayan V, Daniel AK. A novel approach for cluster head selection using trust function in WSN. Scalable Comput Pract Exp. 2021;22(1):1–13.

[10] Narayan V, AK D. CHOP: Maximum Coverage Optimization and Resolve Hole Healing Problem using Sleep and Wake-up Technique for WSN. ADCAIJ Adv Distrib Comput Artif Intell J. 2022;11(2):159–78.

[11] Narayan V, Daniel AK. Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model. 2022;

[12] Srivastava S, Singh PK. Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm. Int J Next-Generation Comput. 2022;13(3).

[13] Srivastava S, Singh PK. HCIP: Hybrid Short Long History Table-based Cache Instruction Prefetcher. Int J Next-Generation Comput. 2022;13(3).

[14] Narayan V, Daniel AK. CHHP: coverage optimization and hole healing protocol using sleep and wake-up concept for wireless sensor network. Int J Syst Assur Eng Manag. 2022;1–11.

[15] Narayan V, Daniel AK. IOT Based Sensor Monitoring System for Smart Complex and Shopping Malls. In: International Conference on Mobile Networks and Management. 2021. p. 344–54.

[16] Narayan V, Mehta RK, Rai M, Gupta A, Singh M, Verma S, et al. E-Commerce recommendation method based on collaborative filtering technology. Int J Curr Eng Technol. 2017;7(3):974–82.

[17] Srivastava S, Sharma S. Analysis of Cyber Related Issues by Implementing Data Mining Algorithm. In: 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence). 2019. p. 606–10.

[18] Smiti P, Srivastava S, Rakesh N. Video and Audio Streaming Issues in Multimedia Application. In: 2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence). 2018. p. 360–5.

[19] Narayan V, Daniel AK. Design Consideration and Issues in Wireless Sensor Network Deployment.”. Invertis J Sci \& Technol. 2020;101.

[20] Narayan V, Mehta RK, Rai M, Gupta A, Tiwari A, Gautam D, et al. To Implement a Web Page using Thread in Java. 2017;

[21] Awasthi S, Srivastava AP, Srivastava S, Narayan V. A Comparative Study of Various CAPTCHA Methods for Securing Web Pages. In: 2019 International Conference on Automation, Computational and Technology Management (ICACTM). 2019. p. 217–23.

[22] Awasthi, S., Chowdhury, B., Metya, S. K., & Majumder, A. (2023). Pockel’s effect inspired Toffoli Gate: an MZI count optimized design and logical applications. Optical and Quantum Electronics, 55(1), 65.

[23] Awasthi, S., Chowdhury, B., Jalil, M. A., Ali, J., Yupapin, P., Metya, S. K., & Majumder, A. (2022). Analysis for Cost Optimized EO Design of a Reversible Boolean Function Using MZIs. In Optical and Wireless Technologies: Proceedings of OWT 2021 (pp. 45-55). Singapore: Springer Nature Singapore.

[24] Awasthi, S., Travieso-González, C. M., Sanyal, G., & Singh, D. K. (Eds.). (2021). Artificial intelligence for a sustainable industry 4.0. Springer International Publishing.

[25] Awasthi, S., Kumar, N., & Srivastava, P. K. (2021). An epidemic model to analyze the dynamics of malware propagation in rechargeable wireless sensor network. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1529-1543.

[26] Awasthi, S., Kumar, N., & Srivastava, P. K. (2020). A study of epidemic approach for worm propagation in wireless sensor network. In Intelligent Computing in Engineering: Select Proceedings of RICE 2019 (pp. 315-326). Springer Singapore