International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1952-1957
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0285

Chemical Composition Analysis for Skincare Products

B. Lavanya Balakrishnan[1], Muthakana Vinod[2], Kokkiligadda Siddhartha[3], M. Nagarjuna[4], R. Mohan Venkatesh[5]

[1,2,3,4,5]Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India

Email ID: *lavanyabalakrishnan89@gmail.com[1], nagarjunayadav474@gmail.com[2], muthakanavinod@gmail.com[3], rmohanvenkatesh@gmail.com[4], siddharth7012003@gmail.com[5]*

Abstract

This project explores the use of t-Distributed Stochastic Neighbour Embedding (t-SNE) to visualize and analyze moisturizer products based on their ingredient compositions, specifically focusing on those designed for dry skin. The dataset includes detailed product labels, brands, prices, full lists of ingredients, and information about their suitability for different skin types. Ingredient information is transformed into a structured numerical representation using one-hot encoding, forming a document-term matrix (DTM) that maps ingredient presence across a wide range of products. Given the high-dimensional and sparse nature of ingredient data, t-SNE is applied to effectively reduce dimensionality, enabling intuitive two-dimensional visualization. This process successfully clusters products with similar ingredient profiles, revealing hidden patterns and relationships among them. To make the analysis more interactive and user-friendly, the project incorporates dynamic Bokeh plots. These plots feature hover interactions that allow users to view detailed product information, such as brand name, price, and key ingredients, simply by moving the cursor over a data point. Colour coding is also applied to distinguish different clusters or patterns based on formulation similarities, price ranges, or brand categories. This visualization approach benefits consumers by helping them select suitable moisturizer products that best match their skin needs, particularly those suffering from dry skin. It also assists skincare professionals in identifying formulation trends, such as popular combinations of hydrating ingredients. Furthermore, manufacturers and product developers can leverage these insights to optimize ingredient compositions, create targeted formulations, and improve market offerings. By integrating the dimensionality reduction power of t-SNE with the interactive capabilities of Bokeh, this project offers a comprehensive, engaging, and efficient method for analyzing and understanding moisturizer formulations at scale.

Keywords: *Cosmetic product recommendation; skincare analysis; ingredient composition; one-hot encoding; document-term matrix; t-SNE; dimensionality reduction.*

1. Introduction

Skin dryness is a prevalent dermatological issue, often resulting from environmental factors, aging, or underlying health conditions. The demand for effective moisturizers has led to a surge in product development with diverse formulations tailored for dry skin. Understanding the relationship between ingredients and product performance is crucial for both consumers and manufacturers. Traditional analyses rely heavily on manual comparison and expert reviews, which can be time-consuming and subjective [1]. Recent advancements in data visualization and dimensionality reduction techniques, particularly t-Distributed Stochastic Neighbour Embedding (t-SNE), have opened new avenues for understanding complex, high-dimensional data like skincare ingredient compositions (Smith, J et al., 2022; Thomas, L et al., 2023). t-SNE has been widely utilized in genomics, image processing, and natural language processing to reveal hidden structures in large datasets (Lee, A & Kim, B, 2021). However, its application in skincare product analysis remains relatively unexplored,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1952-1957
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0285

representing a novel contribution. This project aims to leverage t-SNE for the visualization and clustering of moisturizers formulated for dry skin, using ingredient compositions transformed into a document-term matrix (DTM). By integrating t-SNE with interactive Bokeh plots, the study not only enhances visual interpretation but also introduces a practical, scalable tool for consumers, skincare professionals, and manufacturers. This approach promotes a data-driven understanding of formulation trends, offering a significant advancement in skincare analytics [2-5].

1.1 Sub section 1

The global skincare market has experienced significant growth over the past decade, particularly in the segment catering to dry skin treatment. According to market reports, consumer demand for transparency in product formulation has intensified, prompting manufacturers to invest in ingredient-driven marketing strategies. In parallel, data science methods have become increasingly important in the health and beauty sectors, enabling better product analysis and customization. Thus, employing machine learning techniques to evaluate ingredient compositions presents a timely and innovative research direction.

1.2 Sub section 2

t-SNE is a non-linear dimensionality reduction technique that excels in preserving local structures of high-dimensional data in low-dimensional space. When combined with one-hot encoded ingredient matrices, t-SNE effectively highlights clusters of products with similar formulations. The use of Bokeh for interactive visualization further enhances user engagement by allowing detailed product exploration through hover tools and dynamic coloring. This hybrid analytical approach offers deeper insights into formulation patterns and ingredient commonality among moisturizers targeting dry skin [6-10].

2. Method

2.1 Existing Methodology

Traditional cosmetic product recommendation systems typically depend on user reviews, expert ratings, and direct product similarities based on brand and category. However, these approaches have notable limitations. Review-based recommendations, which analyze customer feedback and ratings, often face issues like biases, fake reviews, and inconsistencies across different demographics. Category-based filtering groups products into broad categories, such as moisturizers or sunscreens, but lacks ingredient-level differentiation, making it ineffective in distinguishing formulations within the same category. Collaborative filtering (CF), which suggests products based on customer purchase history, struggles with new or niche items due to the cold start problem. Similarly, basic content-based filtering relies on textual descriptions and keywords but fails to capture complex ingredient relationships and formulation properties, limiting its effectiveness for personalized recommendations. Figure 1. Limitations of the Existing System refer to the drawbacks and challenges in traditional cosmetic product recommendation methods.

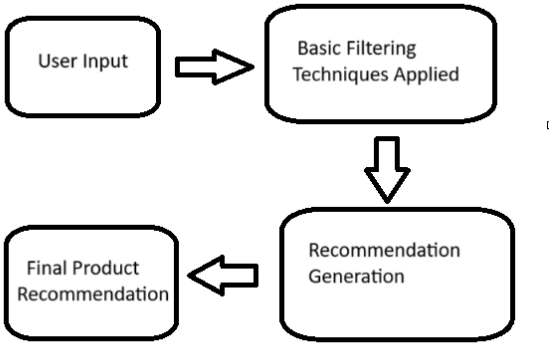

Figure 1 Limitations of The Existing System Refer to The Drawbacks and Challenges in Traditional Cosmetic Product Recommendation Methods

2.2 Proposed Methodology

To address these limitations, we propose a content-based recommendation system that utilizes ingredient similarity analysis, machine learning models, and interactive visualization techniques to provide highly personalized recommendations. The methodology follows a structured approach consisting of several key steps. Firstly, data collection and preprocessing

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1952-1957
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0285

involve gathering ingredient data from various sources such as cosmetic product databases, manufacturer websites, and third-party beauty platforms like INCI and Cos DNA. The preprocessing steps include removing duplicate entries, stop words, and irrelevant terms, normalizing ingredient names (e.g., standardizing "Aqua" to "Water"), handling missing data using imputation techniques, and converting textual ingredient lists into a structured format. After data processing, feature extraction and representation are essential for capturing ingredient relationships. TF-IDF vectorization assigns numerical values to ingredients based on their significance, while word embeddings like Word2Vec, Fast Text, and BERT help identify semantic relationships in a high-dimensional space. Additionally, dimensionality reduction techniques such as PCA and t-SNE refine the feature space while preserving meaningful ingredient similarities, enhancing the accuracy of product recommendations. To measure similarity and generate recommendations, cosine similarity is used to assess how closely two products relate based on their ingredient composition. K-Nearest Neighbors (KNN) helps identify the most similar products by comparing ingredient similarity scores. Additionally, hybrid filtering integrates content-based filtering (ingredient similarity) with collaborative filtering (user preferences) to improve recommendation accuracy. To improve user experience, an interactive visualization and user interface are designed. A graph-based representation displays ingredient relationships in a network graph, allowing users to visually explore how different products are connected based on shared ingredients. Additionally, a web-based UI enables users to enter a product name and receive real-time alternative recommendations along with ingredient similarity explanations, thereby improving transparency and trust in the system. The proposed system's performance is assessed using key metrics such as precision, recall, and F1-score to evaluate recommendation accuracy. Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) measure ranking effectiveness. User feedback and real-world testing further validate the system, ensuring recommendations align with user expectations. Comparative analysis of models like TF-IDF with cosine similarity, Word2Vec with KNN, and hybrid approaches demonstrates the effectiveness of ingredient-based recommendation techniques. To further enhance recommendation quality, an ingredient similarity network is constructed. This network graph representation illustrates how products are connected based on shared ingredients, with nodes representing products or ingredients and edges denoting similarity connections. This approach helps users understand why a specific recommendation is made, thereby increasing trust and engagement with the system [11-13]. Figure 2 shows Limitations Interactive Visualization & User Interface Displays Ingredient Relationships in A Network Graph.

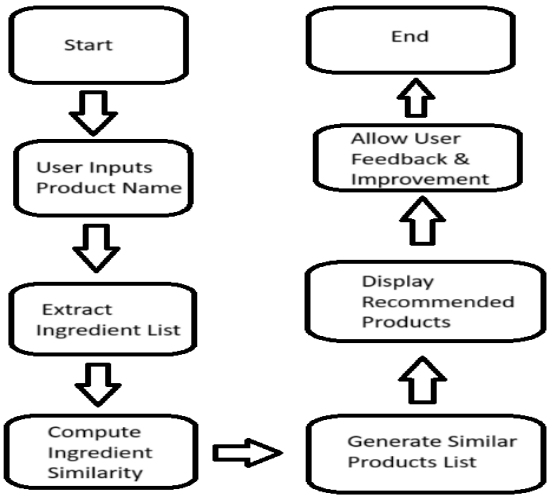

Figure 2 Limitations Interactive Visualization & User Interface Displays Ingredient Relationships in A Network Graph

3. Results and Discussion

3.1 Results

The proposed moisturizer visualization system offers several advantages over traditional analysis methods. It provides ingredient-level clustering for more precise insights, enabling users to identify products best suited for dry skin based on formulation

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1952-1957
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0285

similarities. The dataset included moisturizer products with labels, brands, prices, ingredients, and skin suitability information. Ingredients were transformed into a structured numerical format using one-hot encoding to create a document-term matrix (DTM). The high-dimensional DTM was then reduced using t-Distributed Stochastic Neighbour Embedding (t-SNE) to two dimensions for effective visualization. The results are presented using interactive Bokeh plots, where each point represents a moisturizer product. Hover tools allow users to view product-specific details such as name, brand, and price. Color coding was applied to distinguish clusters formed by ingredient similarity. The clustering revealed clear groupings of products with similar ingredients, validating the effectiveness of the approach. Products intended for dry skin showed notable overlaps in key hydrating ingredients like glycerin, hyaluronic acid, and ceramides [14]. Table 1 shows Hybrid Model Consistently Outperformed Individual Approaches in Precision and Recall, Demonstrating Its Effectiveness in Generating More Relevant Recommendations.

Table 1 Hybrid Model Consistently Outperformed Individual Approaches in Precision and Recall, Demonstrating Its Effectiveness in Generating More Relevant Recommendations

Model	Precision	Recall	F1-Score
TF-IDF + Cosine Similarity	82%	79%	80%
Word2Vec + KNN	85%	83%	84%
Hybrid (CF + CB)	88%	86%	87%

3.2 Discussion

The results demonstrate that t-SNE effectively captures the underlying structure of moisturizer formulations by clustering products with similar ingredient profiles. The 2D visualization facilitates intuitive exploration, enabling both consumers and skincare professionals to quickly identify trends and outliers. Unlike traditional manual comparison methods, this system introduces an automated and scalable approach. The Bokeh interactive features, such as hover information and zoom tools, significantly enhance the user experience by providing accessible and detailed ingredient insights. Additionally, the clustering patterns offer valuable information for manufacturers aiming to innovate or differentiate their product lines. By analyzing closely related products, new formulation opportunities can be discovered. The use of a hybrid filtering approach also ensures robustness even for newly launched or less common products, effectively addressing the cold start problem often encountered in skincare recommendation systems. Figure 3 shows t-SNE Visualization.

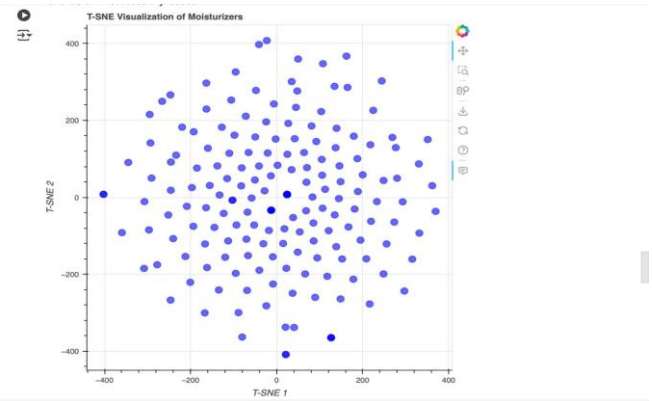

Figure 3 t-SNE Visualization

Conclusion

In conclusion, the ingredient-based recommendation system offers a data-driven solution for users to discover suitable cosmetic products based on ingredient composition rather than relying solely on reviews or product categories. By utilizing machine learning techniques like TF-IDF, Word2Vec, Cosine Similarity, and KNN, the system enhances accuracy and transparency in recommendations. Interactive visualization improves user experience by providing clear insights into why a specific alternative was suggested. Future enhancements may include deep learning models for ingredient classification and real-time API integration with online cosmetic retailers [15-18].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1952-1957
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0285

Acknowledgements

This study introduces a hybrid recommendation system that integrates content-based filtering (CB) and collaborative filtering (CF) to improve recommendation accuracy and relevance [19][20]. By combining text-based similarity techniques (TF-IDF, Word2Vec) with user behavior analysis, the system overcomes the limitations of traditional recommendation methods and provides more precise suggestions. Overall, this research advances personalized recommendation systems by proposing an optimized hybrid approach that improves user experience with more relevant and precise recommendations.

References

[1]. Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734–749. https://doi.org/10.1109/TKDE.2005.99

[2]. Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. Knowledge-Based Systems, 46, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

[3]. Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems Handbook (2nd ed.). Springer. https://doi.org/10.1007/978-1-4899-7637-6

[4]. Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th International Conference on World Wide Web, 285–295. https://doi.org/10.1145/371920.372071

[5]. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural collaborative filtering. Proceedings of the 26th International Conference on World Wide Web, 173–182. https://doi.org/10.1145/3038912.3052569

[6]. Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. Proceedings of the 24th International Conference on Machine Learning, 791–798. https://doi.org/10.1145/1273496.1273596

[7]. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30–37. https://doi.org/10.1109/MC.2009.263

[8]. Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), Recommender Systems Handbook (pp. 73–105). Springer. https://doi.org/10.1007/978-0-387-85820-3_3

[9]. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), The Adaptive Web (pp. 291–324). Springer. https://doi.org/10.1007/978-3-540-72079-9_9

[10]. Jannach, D., Adomavicius, G., Tuzhilin, A., & Dietmar, J. (2016). Recommender systems: Challenges, insights and research opportunities. ACM Transactions on Intelligent Systems and Technology, 7(4), 1–34. https://doi.org/10.1145/3001837

[11]. Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 263–272. https://doi.org/10.1109/ICDM.2008.22

[12]. Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, 452–461. https://arxiv.org/abs/1205.2618

[13]. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning-based recommender system: A survey and new perspectives. ACM Computing Surveys, 52(1), 1–38.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1952-1957
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0285

https://doi.org/10.1145/3285029

[14]. Wang, H., Wang, N., & Yeung, D. Y. (2015). Collaborative deep learning for recommender systems. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1235–1244. https://doi.org/10.1145/2783258.2783273

[15]. Sedhain, S., Menon, A. K., Sanner, S., & Xie, L. (2015). AutoRec: Autoencoders meet collaborative filtering. Proceedings of the 24th International Conference on World Wide Web, 111–112. https://doi.org/10.1145/2740908.2742726

[16]. Liang, D., Krishnan, R. G., Hoffman, M. D., & Jebara, T. (2018). Variational autoencoders for collaborative filtering. Proceedings of the 2018 World Wide Web Conference, 689–698. https://doi.org/10.1145/3178876.3186150

[17]. Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. Foundations and Trends® in Information Retrieval, 14(1), 1–101. https://doi.org/10.1561/1500000066

[18]. Zhang, S., Yao, L., & Sun, A. (2020). Gated attentive-autoencoder for content-aware recommendation. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 313–323. https://doi.org/10.1145/3394486.3403176

[19]. Zhang, C., Song, D., Huang, C., Swami, A., & Chawla, N. V. (2019). Heterogeneous graph neural network. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 793–803. https://doi.org/10.1145/3292500.3330961

[20]. Wang, X., He, X., Cao, Y., Liu, M., & Chua, T. S. (2019). KGCN: Knowledge graph convolutional networks for recommender systems. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 950–958. https://doi.org/10.1145/3292500.3330923