Toxicity Analyzer of Cosmetic Products and Recommendation System

Nishanth[1], Sayeel Shetty[2], Nidhin Shetty[3], Varsha J[4], Spoorthi. B[5]
[1,2,3,4]*Students, Department of Computer Science and Engineering, Srinivas Institute of Technology (S.I.T), Mangalore, Karnataka, India*
[5]*Assistant Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology (S.I.T), Mangalore, Karnataka, India*
doi.org/10.64643/IJIRTV12I12-206777-459

***Abstract*—Cosmetic products are widely used in daily life for skincare, haircare, and hygiene, yet many contain harmful ingredients that can cause health issues such as skin irritation, hormonal imbalance, or long-term diseases. Identifying these toxic components is challenging due to complex chemical names on product labels. To address this, the Toxicity Analyzer of Cosmetic Products and Recommendation System is developed as a machine learning–based solution that analyzes ingredients and evaluates their toxicity levels. Using TF-IDF vectorization, the system compares products and suggests safer alternatives. Implemented in Python with Pandas, NumPy, Scikit-learn, and Gradio, it provides an intelligent and user-friendly approach to help consumers make safer and more informed decisions.**

***Index Terms*—Cosmetic Products, Machine Learning, Recommendation System, TF-IDF, Toxicity Analysis**

I. INTRODUCTION

In today's world, cosmetic products are widely used for skincare, haircare, and personal hygiene. However, many products contain harmful ingredients that may cause skin irritation, allergies, or long-term health issues. Identifying these toxic components is difficult due to complex chemical names on product labels. To address this issue, the Toxicity Analyzer of Cosmetic Products and Recommendation System is developed. It is a machine learning–based system that analyzes ingredient lists, evaluates toxicity levels, and classifies products as Safe, Moderate, or Toxic. The system uses TF-IDF vectorization to compare products and recommend safer alternatives. It is implemented using Python, with Pandas and NumPy for data processing, Scikit-learn for analysis, and Gradio for an interactive user interface.

II. PROPOSED SYSTEM

The proposed system is a machine learning–based application designed to analyze cosmetic product ingredients and evaluate their toxicity levels. The system allows users to select a product and specify their skin type, after which it processes the ingredient list and identifies harmful chemicals using a predefined toxicity database.A toxicity score is calculated based on the presence of harmful ingredients, and the product is classified as Safe, Moderate, or Toxic. To assist users in choosing better products, the system uses TF-IDF vectorization and cosine similarity to compare products and recommend safer alternatives with similar ingredient compositions. The system is implemented using Python, with Pandas and NumPy for data processing, Scikit-learn for machine learning operations, and Gradio for a simple and interactive user interface.

III. METHODOLOGY

A. Data Collection and Preprocessing

The system uses a cosmetic product dataset containing product name, category, and ingredient list. The dataset is loaded using the Pandas library in Python. Preprocessing is performed to improve data quality by removing duplicate entries, converting all ingredient names to lowercase, and eliminating unnecessary spaces and special characters. Ingredient lists are also tokenized into individual components to enable accurate analysis. This step ensures that the data is clean, consistent, and suitable for further processing.

B. Ingredient Toxicity Analysis
The preprocessed ingredient data is analyzed using a predefined toxicity database that contains information about harmful chemicals and their effects on different skin types such as normal, oily, dry, and sensitive. Each ingredient in the selected product is compared with this database to identify toxic components. Based on the number and severity of harmful ingredients detected, a toxicity score is calculated.

C. Recommendation Process
To provide safer alternatives, the system uses TF-IDF (Term Frequency–Inverse Document Frequency) vectorization to convert ingredient text into numerical feature vectors. Cosine similarity is applied to measure the similarity between the selected product and other products in the dataset. Products that belong to the same category and have lower toxicity scores are filtered and ranked based on similarity. The top similar and safer products are then recommended to the user.

D. Result Visualization
The final results are presented through an interactive user interface developed using Gradio. The system displays product details, ingredient lists, detected harmful ingredients, and a toxicity score with classification. Additionally, recommended safer alternatives are shown along with similarity scores. This visualization helps users easily interpret the results and make informed decisions.

IV. SYSTEM DESIGN

Figure. 1 shows the system architecture of the Toxicity Analyzer of Cosmetic Products and Recommendation System. The process begins with user input, where the product and skin type are selected. The input data is then processed through the data preprocessing module, which cleans and standardizes the ingredient information. The ingredient analysis module identifies harmful ingredients by comparing them with a predefined toxicity list. Based on this analysis, the system calculates a toxicity score and classifies the product as Safe, Moderate, or Toxic. The results are then displayed to the user in an understandable format. Further, the recommendation module uses TF-IDF and cosine similarity to compare products and suggest safer alternatives with similar ingredient compositions. This architecture ensures accurate toxicity analysis and helps users make informed decisions.

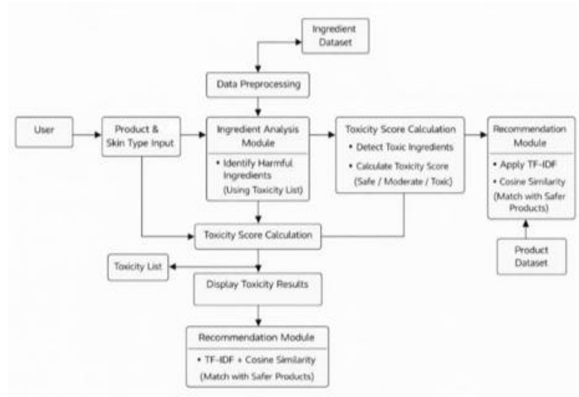

Figure. 1: System Architecture of Toxicity Analyzer

V. RESULTS

The developed system successfully analyses cosmetic products based on their ingredient composition and provides toxicity evaluation along with safer product recommendations. The user selects a product and skin type through the interface, after which the system processes the ingredient list and identifies harmful chemicals using a predefined toxicity database. The system calculates a toxicity score and classifies the product as Safe, Moderate, or Toxic. For example, products with fewer harmful ingredients are classified as safe, while those with higher toxicity are categorized as moderate or toxic. The results are displayed clearly, highlighting harmful ingredients and toxicity levels.In addition, the system recommends safer alternative products using TF-IDF vectorization and cosine similarity. These recommendations are based on ingredient similarity and lower toxicity scores. The results demonstrate that the system effectively helps users identify safe cosmetic products and make informed decision.

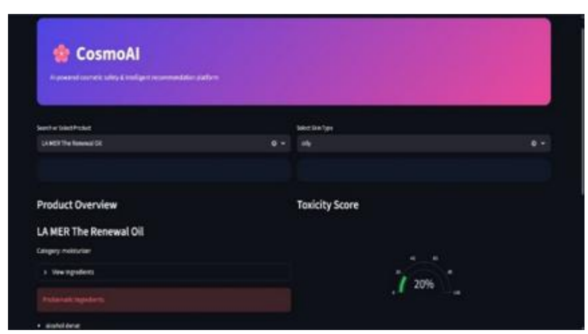

Figure 2: Safety Analysis Interface

Figure. 2 shows the output of the system when a cosmetic product and skin type are selected. The product “LA MER The Renewal Oil” with oily skin type is analyzed, and the system detects Alcohol Denat as a harmful ingredient. A toxicity score of 20% is calculated, and the product is classified as Safe. The result is displayed using a gauge meter along with highlighted harmful ingredients for better understanding.

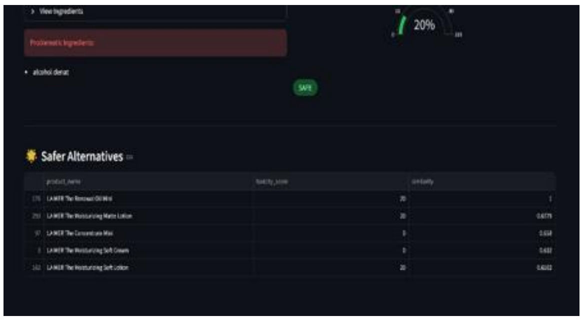

Figure 3: Alternative Product Recommendations

Figure. 3 shows the safer product recommendations generated by the system. Using TF-IDF and cosine similarity, the system identifies products with similar ingredients but lower toxicity scores. It recommends alternatives such as LA MER The Renewal Oil Mini, The Moisturizing Matte Lotion, and The Concentrate Mini. Each product is displayed with its toxicity score and similarity value, helping users choose safer options.

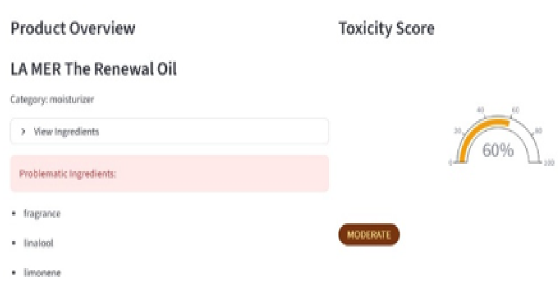

Figure 4: Moderate Analysis Interface

In Figure 4, once the product and skin type are selected, the system retrieves the ingredient list from the dataset and performs toxicity analysis using the predefined toxicity database. During the ingredient evaluation process, the system identifies fragrance, linalool, and limonene as problematic ingredients for sensitive skin. These ingredients are known to cause irritation or allergic reactions in sensitive skin types.

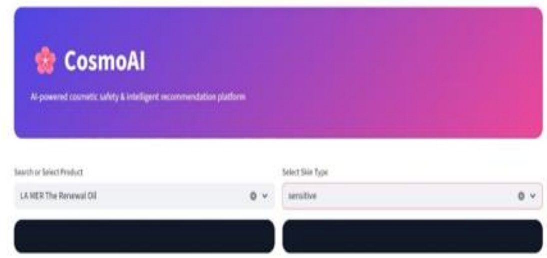

Figure 5: Moderate Toxicity Detection score

Figure. 5 shows that the calculated score falls within the moderate toxicity range, and therefore the product is classified as MODERATE. The toxicity score is visually represented using a gauge meter, which helps users easily understand the safety level of the selected cosmetic product. After analyzing the product, the system generates safer alternative recommendations. The recommendation system uses TF-IDF vectorization to convert ingredient lists into numerical vectors and applies cosine similarity to measure the similarity between the selected product and other products in the dataset.

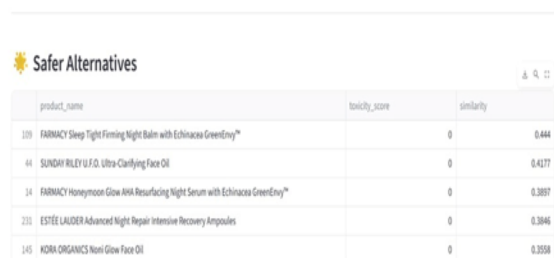

Figure 6: Recommended Safety Products

Figure. 6 shows that the system recommends safer cosmetic products based on ingredient similarity and lower toxicity scores. Each recommended product is displayed with its toxicity score and similarity value, enabling users to compare different options. The system also ensures that all recommended products belong to the same category as the selected product, maintaining functional relevance. This feature helps users identify safer alternatives and make informed decisions based on their preferences and safety requirements.

VI. CONCLUSION

The Toxicity Analyzer of Cosmetic Products and Recommendation System demonstrate the effective use of machine learning techniques to evaluate the safety of cosmetic products. The system successfully analyzes ingredient lists, identifies harmful components, and classifies products based on their toxicity levels. By incorporating TF-IDF vectorization and cosine similarity, it also provides meaningful recommendations for safer alternatives within the same category.

The developed system is user-friendly, efficient, and capable of assisting consumers in making informed decisions about cosmetic product usage. It reduces the difficulty of understanding complex ingredient labels and promotes awareness about product safety. Overall, the project highlights the potential of integrating data analysis and recommendation systems to enhance consumer health and safety.

VII. APPENDIX

The Toxicity Analyzer system allows users to select a cosmetic product and input their skin type through a simple interface. The system processes the ingredient list and compares it with a predefined toxicity database to identify harmful substances. Based on the detected ingredients, a toxicity score is generated and the product is classified as Safe, Moderate, or Toxic.

In addition, the system provides recommendations for safer alternatives. It uses TF-IDF vectorization and cosine similarity to compare products and suggest those with similar ingredients but lower toxicity levels. The final output includes toxicity details, highlighted harmful ingredients, and recommended products, helping users make safer and more informed decisions.

The system is implemented using Python and integrates libraries such as Pandas and NumPy for data preprocessing, Scikit-learn for feature extraction and similarity computation, and Gradio for building an interactive user interface. The overall workflow is designed to be efficient and user-friendly, ensuring quick analysis and easy interpretation of results. This structured approach enhances the reliability of toxicity evaluation and improves the overall user experience.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to their project guide, Prof. Spoorthi B, for her valuable guidance, support, and encouragement throughout the project. We also thank the Head of the Department and faculty members of the Computer Science and Engineering department for their support.

REFERENCES

[1] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2021.

[2] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson Education, 2023.

[3] Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2022.

[4] C. Aggarwal, *Recommender Systems: The Textbook*. Springer, 2016.

[5] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*. Springer, 2015.

[6] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.

[7] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," in *Proceedings of the First Instructional Conference on Machine Learning*, 2003.

[8] G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[9] P. D. Turney and P. Pantel, "From Frequency to Meaning: Vector Space Models of Semantics," *Journal of Artificial Intelligence Research*, 2010.

[10] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, 2003.

[11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2011.

[12] F. Pedregosa *et al*., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[13] W. McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Python*. O'Reilly Media, 2018.

[14] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, 2007.

[15] F. Chollet, *Deep Learning with Python*. Manning Publications, 2018.

[16] J. Bae and M. Piao, "Cosmetic Ingredient Analysis and Safety Evaluation Using Machine Learning Techniques," *International Journal of Cosmetic Science*, 2020.

[17] European Commission, *CosIng Database: Cosmetic Ingredient Database*, 2019.

[18] U.S. Food and Drug Administration, *Cosmetic Ingredient Safety Guidelines*, 2022.

[19] Organisation for Economic Co-operation and Development (OECD), *Guidelines for the Testing of Chemicals and Cosmetic Ingredient Safety*, 2018.

[20] W. McKinney and Pandas Development Team, *Pandas: Data Analysis and Manipulation Library Documentation*, 2023.

[21] F. Pedregosa *et al*., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[22] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2013.

[23] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[24] J. H. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. Springer, 2001.