Personalized Smart Skincare Product Recommendation System

L.M.I.T. Hemantha
Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka

T.M.E. Gayathri
Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka

N.N.M. De Silva
Department of Information Technology Sri Lanka Institute of Information Technology Malabe, Sri Lanka

ABSTRACT
Cosmetics have been part of the lifestyles of humans since civilization began, and globalization has driven the cosmetics industry to develop several new solutions. Chemical stability is a key component of ensuring that a cosmetic product is safe for consumer use. Before using a cosmetic product, it is highly recommended to understand and recognize the facial condition of the consumer and give them a clear understanding of the products that they are using. Therefore, the main objective of this study is to recognize the various skin types of consumers and give a realization about the products to prevent Allergies.This component investigates how may derive feature-level evaluations of beauty items from consumer reviews and ratings to impact future consumer and manufacturer decision-making. In the study, user reviews and ratings In this studycollected on online shopping sites for a selection of cosmetic products. For this goal, In this study explicitly did a feature-level and sentiment analysis for reviews.Sentiment analysis is research that analyses and extracts opinions from given reviews. This studyhas shown each review's positive and negative aspects in this section. Finally, myn this researchresearch provides ratings for four cosmetics features. It aids in creating individualized purchasing decisions. In this case, the intended alternative was cosmetics that are selected based on predetermined criteria. This decision support system can provide alternative choices of cosmetics that can be used later as a reference to determine cosmetics that are suitable for the type of facial skin.

General Terms
Image processing, Machine Learning, Artificial Intelligence, and Predictive methods.

Keywords
Image processing, Natural language processing, Sentiment Analysis Introduction, ingredients analysis

1. INTRODUCTION
The community has a very great demand for cosmetics today, especially since most individuals want to seem appealing and stunning all the time and grab people's attention. The idea of human attractiveness is frequently associated with flawless facial skin. Therefore, it is not unexpected that any product on the market that may lighten facial skin is doing in this study. Finding appropriate cosmetics for a certain type of face skin is a difficult process since, in the absence of the use of cosmetics, it is impossible to predict the effects that the skin will have. Additionally, since every person's facial skin is unique and requires a unique approach to treatment, there is not a single face-whitening cosmetic that can be applied to all types of facial skin. A Fuzzy logic technique is used to choose the cosmetics most suited to the kind of facial skin while building a decision support system application. People require proper supervision while using cosmetics available in the market because there is a possibility of customers getting into difficulties if they use these goods without knowledge. These kinds of concerns can sometimes lead to more serious health problems. Therefore, determining the customer's skin type and purchasing goods that are suitable, as In this study as learning about the ingredients that In this study used to make these products, will help to reduce the risk. As for the information that In this studyhas gathered from past research, In this studyis hoping to get accurate responses from customers who can bring a significant impact to the productivity of the brand company and enhance the positive image that society has of it.

Customers frequently base their decisions on user reviews left by previous customers. In this study, most of these review on In this study sites are generic and lack specificity. Even if products may be compared using the offered product-level ratings, there will always be a certain percentage of individuals who prefer buying the things based on certain characteristics. Such individuals typically have to read all the comments in order to learn how past users felt about the aspects of the product in question. Finding the best product for themselves becomes a difficult process for a consumer when there are so many options available for a particular item (like a cosmetic). Reviews and comments can be used to identify specific user preferences. By gathering details about individual preferences, recommendations enable customers to make knowledgeable decisions regarding a variety of products. Furthermore, such product-level assessments rarely explain what is excellent or terrible about the product from the viewpoint of the manufacturer. As a result, having feature-level evaluations available gives the producer more knowledge about how to enhance the product. With all these benefits in mind, In this study goal is to develop a feature-level rating system. Additionally, this system will provide users with the choice to offer product feedback based on their preferences.

Cosmetics is any substance used to clean, improve or change the appearance, of skin, nails, hair, teeth, and many more. It includes beauty preparation such as make-up, perfume, skin cream, nail polish, and grooming aids which are known to us as shampoo, soap, deodorant and etc. This research will be centered on recommending healthy and sustainable cosmetic products to customers. When used under usual or foreseeable situations, a cosmetic product that is put on the market must not impair human health or the body in any way. The main motive would be to have a better understanding of the ingredients/chemicals that the cosmetics products are made of and recommend healthy and sustainable ayurvedic/herbal cosmetic products/ingredients to consumers as a substitute for artificial products that contain fair skin.

2. LITERATURE REVIEW

2.1 Skin type identification

Some journal details In this study Cosmetics all around the world [1].From ancient times people used to use advertisements to choose their beauty products. But nowadays due to the wide distribution of beauty salons, consumers are accustomed to seeking the help of people in beauty salons to recommend cosmetics for their skin.[6] [5]This is not a very successful method as the cosmetics used in cosmetics are limited to a certain brand name. This is because beauty salons are often run by companies of different brands [5]. This is not a very successful method as the cosmetics used in cosmetics are limited to a certain brand name. This is because beauty salons are often run by companies of different brands [5].As a result, cosmetic users have been forced to use one-brand cosmetic products that do not suit their skin. As a Side effect of this habit, young people are exposed to a variety of skin allergens [3] [7] At present, online marketing platforms are being used to prescribe cosmetics based on consumer purchases, but it is not a successful method. Because those recommendations are based on other consumers' purchases and their own reviews, those recommendations are made without any skin condition being identified [8]. The best solution is to break consumers' needs or problems into a series of smaller steps. Then choose the cosmetics that suit your skin from a database used in the recommendation system as mentioned above, there are several different media and ways currently used by consumers to recommend cosmetics In this studyproposes to set up a system to check the condition of the skin from several different angles to identify the most successful skin-friendly cosmetics, Identify the user skin color, Identify the physical condition of the skin, Identify the physical condition of the environment around the skin

After studying various experiments conducted by different people, it was found that there was no perfect system to identify all the conditions of the face and to approve the suitable creams. Many systems recommend coatings at the request of other users and their constituents. As mentioned above, the basic part of the system In this studyis going to introduce prescribing cosmetics to identify the physical condition of the skin. For that, it is important to know the skin color in this study as the surface condition of the skin.

2.2 Review Analysis

Customer reviews play an increasingly important role in consumers' purchasing decisions. Sentiment analysis is such a research area that understands and extracts the opinion from the given review and the analysis process includes natural language processing (NLP), computational linguistics, text analytics, and classifying the polarity of the opinion. Sentiment analysis [13],[14],[15],[16] has been an active research topic for a long period now. It has applications in health, politics, sports, e-commerce, and so on. In e-commerce, customer reviews can give lots of insights about the products, as shown in and through sentiment analysis.[14]. The analysis, in this study, is limited to general reviews and not to specific qualities of the cosmetic product. which has never been done in previous studies.

Beauty products largely depend on the opinion of others who have already purchased the product and used it. One of the best ways to watch the review section and see opinions and comments given by other users to gain some knowledge about a product. And when they have rated products, users can easily share their views according to their purpose. And Customers can easily decide whether the product posted is good or bad by using this application. Beauty products reviews and ratings are also more important to get an idea about their purposes.Reviews and ratings can help you quickly come to a good conclusion from someone else's ideas and choose a beauty product that suits you quickly. feature-level ratings of the mobile products from the customer reviews and review votes to influence decision-making, both for new customers and manufacturers. Such a rating system gives a more comprehensive picture of the product than what a product-level rating system offers. While product-level ratings are too generic, feature-level ratings are particular [14]. There when applying features level rating for the cosmetic product it is easy for consumers to easily select their items according to their purpose

2.3 Product type and ingredient analysis

Sometimes a product can be cosmetic as in this study as a drug at the same time. It happens when a product is made for two uses. As an example, an antidandruff shampoo is considered both a cosmetic and a drug. The shampoo will be identified as cosmetic since it is intended to clean the hair, while the antidandruff identifies as a drug because it is intended to treat/cure dandruff. Likewise, in this study can identify the products which are both cosmetic and drugs. So, the proposed component will be identifying whether the product is a drug or cosmetic, or both in order to get a clear understanding before using any product [18][19][20][21].

It is against the law for a cosmetic to contain the above chemicals but sometimes this list may vary from country wise. So, the restricted list will be changed if there is any product that was manufactured in a different country. Skin infections are shared among all age groups and can be due to the introduction of microbes, chemicals, and biological toxins within the environment, and a few expand due to ailing health. They had to depend on the information of nature compiled within the Ayurveda as it was calculated. The science of Ayurveda utilized numerous herbs and vegetation to create beauty care products for beautification and assurance from outside influences. The typical substance within the botanicals does not cause any side impacts on the human body; instead, improves the body with supplements and other valuable minerals.[24]

The place for ayurvedic cosmetic products is developing very quickly. Numerous companies have entered the fragment with branded items in skincare, haircare, cleansers, and fundamental oils. Concern about harmful, destructive chemicals in cosmetic items has expanded customer interest in natural/ayurvedic/herbal beauty care products. Increasingly more brands now include homegrown and botanical ingredients in their products.

3. METHODOLOGY

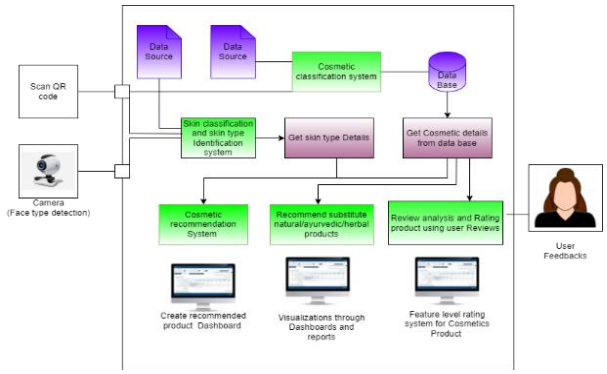


Figure 1. System Overview Diagram

3.1 Skin type identification

OpenCV use for identifying the user's face and extract only cropped images from the user's uploaded images.Face detection using Haar cascades is a machine learning-based approach where a cascade function is trained with a set of input data. OpenCV use for identifying the user's face and extract only cropped images from the user's uploaded image. Then the image obtained by the system is directed to the CNN. The features are extracted from the picture via the convolution technique. The face image thus analyzed by the neural network is processed with the pre-trained model.

According to the information found through this studies, In this study found that other Roman fonts have a unique skin condition, and after analytically classifying them using a branching system, In this studyIn this study able to divide them into four skin types. They are Oily skin, Dry skin, Normal skin, and combination skin. The final skin selection decision is not made solely by using that pre-trained system. For this, the system is asked several questions. To explain the feature, it is a small quiz. The reason for this is that this system does not touch the skin in any way when choosing the skin type, without touching the skin, the system gets data from quizzesto understand the physical condition of the skin and then gives a separate score to the data obtained from them. Happens After that, the output of the CNN system is combined, and the final skin condition is displayed to the user

3.2 Review Analysis

Sentiment analyses and approaches classify reviews based on study records and review this study rating. *Sentiment analysis* is a study that analyzes and derives opinions from given reviews. This paper will go over a sentiment classification machine learning strategy. This technique is similar to classifying the top into both positive and negative emotion categories.

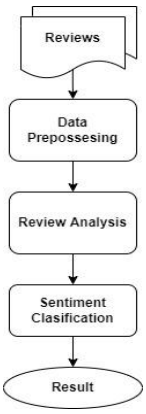


Figure 2. Sentiment Analysis Steps

Review Analysis and sentiment-based feature-level rating systems are important to the customer when taking purchasing decisions. Although this system is not an e-commerce platform, In this studyintends to give users the chance to publish their thoughts and share their experiences using customer product reviews. This research investigates how consumer decisions may be influenced by. consumer reviews. Compared to a rating system at the product level, this one offers a more complete image of the product. Ratings of features are more precise while reviews of products are too generic. figuring out, identifying, or otherwise describing using NLP, statistics, or machine learning approaches to determine the sentiment content of a text unit.

This studyprovides an approach based on an evaluation expression dictionary generated from a corpus of real-world internet reviews for automatic scoring of various features there will be some tools used in hand for backend cosmetic item review texts. This study also thinks about how to measure user similarity in cosmetic user reviews.

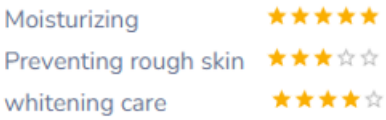


Figure 3. Features Level Ratings

3.3 Product type and ingredient analysis

Predictive analytics is the use of data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on historical data.

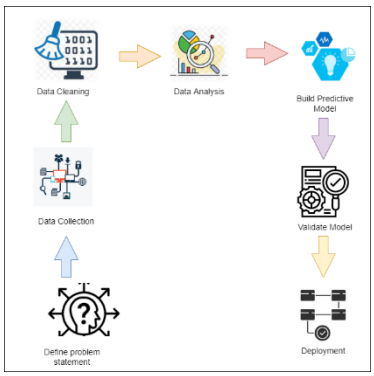


Figure 4. Predictive Analysis Steps

From that using the Jupiter notebook, the predictions in this study done as below,

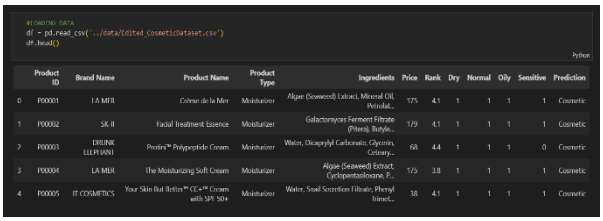


Figure 5 Predicted Result

Analyzed skin types analyzed skin quiz data and product data from different brands would be the data sources that will result in the analysis for this component.

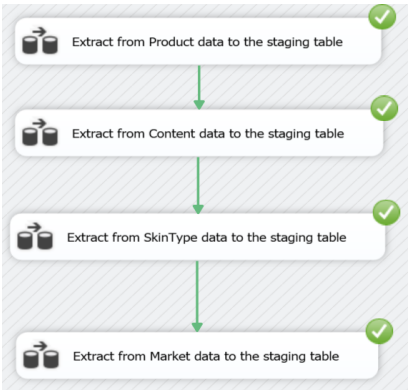


Figure 6. Analyze the process

Extracted data will be pre-processed and then will be transformed into a suitable format that is appropriate to input

to the designed model. The algorithm included in the model. would be implemented by extracting features from classification algorithms that befit such as decision tree, and naïve Bayes algorithms. The data model will be trained and then tested. Then based on the products of the proven factors, substitute products will be recommended using content-based, collaborative filtering. Also, an ETL process will be done, in order to develop dashboards.

Data will be visualized by generating reports and presenting the results using dashboards. These dashboards will be backed up by visualization techniques and tools to ensure that the key performance indicators, metrics, and key results are communicated in an attractive manner. Dashboard summarization will display many important about the products, ingredients and etc. that will give a clear image to the consumer.

4. RESULT AND DISCUSSION

4.1 Skin type identification

To train image data sets a developer must consider the accuracy of the output of the result. Output accuracy depends on the number of epochs of the CNN model. Although this study used the best image processing models to train image datasets. A Python library called OpenCV is mostly used for computer vision and image processing. To find faces in the picture, this studyemploys a Haar Cascade model that has already been trained. The object detection technique used to extract things from the picture is a haar cascade. With the help of many photos, this algorithm was trained.

As mentioned in the methodology, In this studyhas categorized users' skin types as mainly four categories after this literature review [9] [10]. As oily skin, dry skin, normal skin, and Combinational skin. To create a pre-trained model,in this study collect more than thousands of images under details having the above categories. In this study scrape face images under the above categories from Beautiful Soup (bs4) python library. That library pulls data from HTML and XML files. It is an external module for python. For a train,the data sets in this study used CNN (Convolutional Neural network) model. If this studyconsiders its essence a matrix is subjected to a mathematical procedure called convolution. Typically, the picture is represented in this matrix by pixels or integers. To develop that model, as an essential library, this study used NumPy, pandas, and pyplot. When this studydevelops this image processing model, in this study developed an automated epochs identifying system when training the training data set using manual hardcode python function. Developedthe front end in this studyhas used React js and bootstrap.

As discussed, the methodology first identifies the face from user-uploaded images.

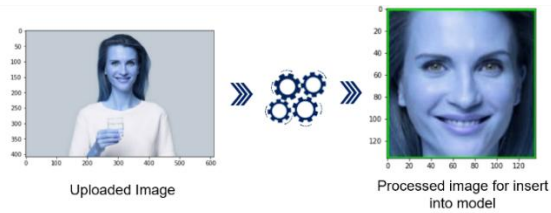


Figure 7.Identificationsof uploaded image

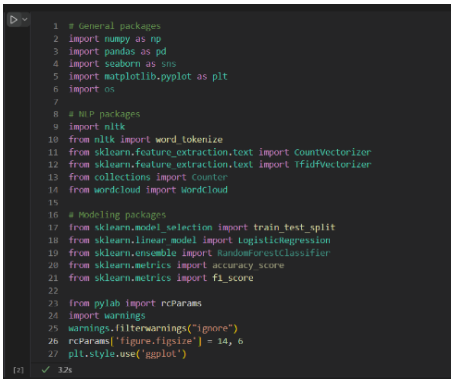
Predicting main keywords to user skin type after identification of all conditions of the skin Oily skin, Dry skin, Combinational skin, Normal skin.

Figure 8. The output of Image Processing

4.2 Review Analysis

This study uses cosmetic customer reviews by using in this study crawling technology. Consumer reviews will be used as a data source in this research. Build an NLP model to Analyze product reviews and understand and derive its opinion from the provided review and analysis Using NLP, statistics, or machine learning approaches, determining, identifying, or otherwise describing the sentiment content of a text unit. This diagram depicts the procedures taken during data processing and analysis. There are numerous phases to the data processing and analysis technique that must be followed In this study. It is usually referred to as "opinion mining," In this study, in this case, the focus is on extraction. The study focuses on rating prediction using social sentiment analysis. The most essential and fundamental addition to sentiment analysis is determining the user's preferences. The most essential and fundamental addition to sentiment analysis is determining the user's preferences. Typically, sentiment analysis seeks to understand a speaker's or author's perspective about a given subject or the polarity of a document's discourse. Reviews are often split into two groups: positive and negative. Customers are more likely to favor purchasing products with excellent evaluations when given a choice and Implement Feature level Ratings for a cosmetic product.

A large amount of effort is focused on component sentiment analysis, including general phase-level sentiment analysis in reviews and papers. As supervised learning machines, In this study may choose Logistic Regression and TF-ID to produce the best results.



```
# General packages
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import os

# NLP packages
import nltk
from nltk import word_tokenize
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from collections import Counter
from wordcloud import WordCloud

# Modeling packages
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import f1_score

from pylab import rcParams
import warnings
warnings.filterwarnings("ignore")
rcParams['figure.figsize'] = 14, 6
plt.style.use('ggplot')
```

Figure 9. Libraries for feature analysis

Reviews classified (Positive & Negative) using the NLP model

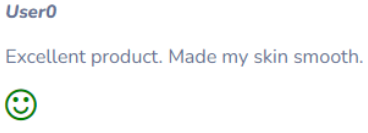


Figure 10. PpositiveReview

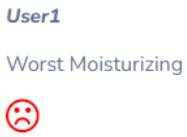


Figure 11. Negative Review

Figure 12. Positive Review Output

When in this study put the new review, it identifies positive or negative and ratings the features according to the review.

Figure 13. Negative Review Output

4.3 Product type and ingredient analysis

All the products and their ingredients will be scraped by the relevant In this study sites and then using Jupyter Notebook which is an open-source In this study application that was used to create and share documents that contain live codes, equations, visualizations, and text to clean, preprocess the data and train them into different algorithms to build models. In order to train data and to get the best accuracy different algorithms such as SVC, Random Forest Classifier, Decision Tree Classifier, Gradient Boost Classifier and etc. are used. To get the predictions python is used as the language and libraries such as NumPy, Scikit-learn, and Pandas are also used and imported to the model building.

VS Code and PyCharm are mainly used to implement the front-end and chatbot. The chatbot will be implemented using the Rasa framework, and it will be done by creating a virtual environment in the workspace mainly in PyCharm. When creating meaningful insights using SQL Server Management Studio (SSMS) which is an integrated environment for managing any SQL infrastructure will mainly be used.

Testing was done by giving an ingredient list to get the prediction, this prediction is getting through the rasa chatbot that was implemented

Testing was done by giving an ingredient list to get the prediction, this prediction is getting through the rasa chatbot that was implemented

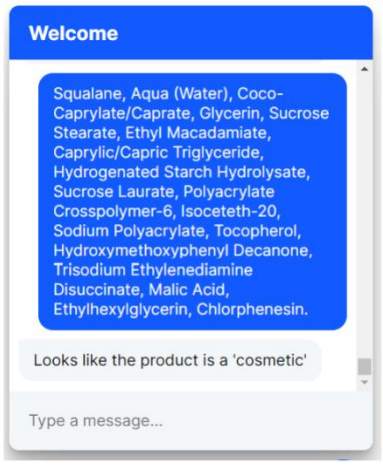


Figure 14. Ingredient list in chatbot

Manual testing was done to see if the predictions are correct. This was cross-checked using the actual data set to whether the predictions are accurate.

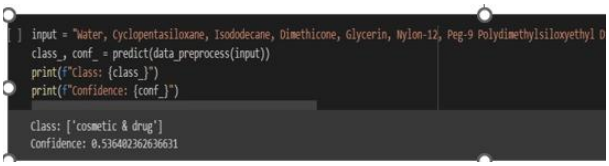


Figure 15. manually predicts the product type

5. CONCLUSION

From this skin classification system hope to develop get most accurate classification systemusing develop CNN model. It can be developed by attaching new layers for softly identifying skin colors and skin differences. from these feature developments, consumers can recommend the most suitable cosmetics by using the output of skin classification. And This studygets 56% accuracy for ratings that have to be predicted. This study will work on improving the performance for features level ratings and in this study can improve the features for ratings.Developing the prediction accuracy to a higher level and trying to predict some other aspects of a product using the ingredients of that particular product. Training the chat bot with many more questions regarding products and ingredients.

6. ACKNOWLEDGEMENT

In this study would first want to express deep gratitude to the lecturer for guiding and supporting us as this study completed this fourth-year project. Without the frequent advice and assistance, this study received, in this study could not have completed this job in such a successful manner. Second, in this study want to thank the lecturer panel for their humility while presenting the concepts and fundamentals to us. Finally, in this study would like to express sincere gratitude to family, friends, and anyone else who helped us, even in a small way, by encouraging and motivating us to finish this project successfully when in this study occasionally felt distracted and discouraged due to the workload in this study had over the past few months.

7. REFERENCES

[1] G. S. G. and M. G. Y. L. Mahagamage, "Selection, usage, and disposal of personal care and cosmetic products in Sri Lanka; an environmental and health concern," repository.kln.ac.lk, 2021.

[2] G. C. P. Dabare and R. a. S. In this studyerasiri, "Factors

Influencing on Purchase Intention for Men's Fairness Cream Products in Sri Lanka.," repository.kln.ac.lk, 2016.

[3] K. Aladin, "(PDF) Journal of Critical Reviews THE IMPACT OF SOCIAL MEDIA ON CONSUMERS' HEALTH BEHAVIOR TOWARDS CHOOSING HERBAL COSMETICS," THE IMPACT OF SOCIAL MEDIA ON CONSUMERS' HEALTH BEHAVIOR TOWARDSCHOOSING HERBAL COSMETICS, 2022.

[4] J.I. Alani, D. Mark Denis P. and Y. James A., "Allergy to Cosmetics: A Literature Review," Dermatitis®, vol. 24, no. 6, p. 283–290, 2013.

[5] B. Magdalena, P. Matei, C. A. Caterina, C. Adriana, S. C. Daniela and D. Stan Cătălina, "Health Risk Assessment of Mercury in Cosmetics," 2021.

[6] P. Ngamjarussrivichai, P. Jeemali and P. Wuttipong, "Designing an online marketing investment evaluation system case study of cosmetics manufacturing company," 2018 5th International Conference on Business and Industrial Research (ICBIR), 2018.

[7] H.-T. Chen and F. Kuo-Kuang, "The study of the factors in the effect of cosmetic advertisement," 2016.

[8] B. Ujkani, D. Minkovska and L. Stoyanova, "A recommender system for WordPress themes using item-based collaborative filtering technique," 2020.

[9] Y. Zhu, C. Huang, K. Zhai and L. Pan, "Face detection method using template feature and skin color feature in rgb color space," 2015.

[10] K. Nallaperumal, S. Ravi, C. N. K. Babu, R. Selvakumar, A. L. Fred and C. Seldev, "Skin Detection Using Color Pixel Classification with Application to Face Detection: A Comparative Study," 2007.

[11] L. Regina and A. Mudrik, "A comparison of human skin color detection for biometrie identification," 2017.

[12] P. Tejal, D. Nishan, J. Amisha, G. Umesh, K. Desai and R. Bansal, "Cosmetics and health: usage, perceptions and awareness," Bangladesh Journal of Medical Science, vol. 12, no. 4, pp. 392-397, 2013.

[13] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," in 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 10-11 March 2017.

[14] K. R. Jerripothula, A. Rai, K. Garg and Y. S. Rautela, "Feature-Level Rating System Using Customer Reviews and Review Votes," in IEEE, 31 July 2020.

[15] C. C. P. Hapsari, W. Astuti and M. D. Purbolaksono, "Naive Bayes Classifier and Word2Vec for Sentiment Analysis on Bahasa Indonesia Cosmetic Product Reviews," in IEEE, Bandung, Indonesia, 06-07 October 2021.

[16] S. N. Singh, "Sentiment Analysis of a Product based on User Reviews using Random Forests Algorithm," in 2020 10th. International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 29-31 January 2020.