LEVERAGING COSMETIC INGREDIENT PROFILES AND SKIN-RELATED FEATURES FOR COSMETIC PRODUCTS RECOMMENDATION

THEINT ZAR LWIN KYAW

MASTER OF SCIENCE

IN

DIGITAL TRANSFORMATION TECHNOLOGY

SCHOOL OF APPLIED DIGITAL TECHNOLOGY

MAE FAH LUANG UNIVERSITY

2024

LEVERAGING COSMETIC INGREDIENT PROFILES AND SKIN-RELATED FEATURES FOR COSMETIC PRODUCTS RECOMMENDATION

THEINT ZAR LWIN KYAW

THIS THESIS IS A PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN DIGITAL TRANSFORMATION TECHNOLOGY

SCHOOL OF APPLIED DIGITAL TECHNOLOGY
MAE FAH LUANG UNIVERSITY
2024

THESIS APPROVAL

MAE FAH LUANG UNIVERSITY

FOR

MASTER OF SCIENCE IN DIGITAL TRANSFORMATION TECHNOLOGY

Thesis Title: Leveraging Cosmetic Ingredient Profiles and Skin-Related Features for Cosmetic Products Recommendation

Author: Theint Zar Lwin Kyaw

Examination Committee:

Associate Professor Natthakan Iam-On, Ph. D.	Chairperson
Assistant Professor Surapong Uttama, Ph. D.	Member
Patcharaporn Panwong, Ph. D.	Member
Assistant Professor Worasak Rueangsirarak, Ph. D.	Member

Advisors:

S Ma
....Advisor
(Assistant Professor Surapong Uttama, Ph. D.)

Patcharaporn
....Co-Advisor
(Patcharaporn Panwong, Ph. D.)

Dean:

....
(Assistant Professor Nacha Chondamrongkul, Ph. D.)

ACKNOWLEDGEMENTS

To everyone who encouraged me throughout my journey to completing this thesis, I would like to extend my sincere gratitude. Most importantly, I am deeply grateful to my advisor, Dr. Surapong Uttama, for invaluable support, encouragement, and guidance. I am also appreciative of my co-advisor, Dr. Patcharaporn Panwong, whose insightful feedback and unwavering assistance have greatly enriched this study. My heartfelt appreciation extends to Asst. Prof. Dr. Worasak Rueangsirarak and Asst. Prof. Dr. Santichai Wicha for sharing their knowledge and expertise during my academic journey. Additionally, I am deeply indebted to Mae Fah Luang University for providing grant support. I also would like to thank the School of Applied Digital Technology for the support resources, facilities, and academic environment essential for completing this study.

My greatest gratitude goes to my family, whose unwavering love, patience, and sacrifices have provided a continual source of inspiration every step of this journey. I also appreciate my friends, seniors, and juniors' support, advice, and camaraderie, which have enhanced this experience. Finally, special thanks to everyone who assisted in the data labeling process.

Theint Zar Lwin Kyaw

Thesis Title	Leveraging Cosmetic Ingredient Profiles and Skin-Related Features for Cosmetic Products Recommendation
Author	Theint Zar Lwin Kyaw
Degree	Master of Science (Digital Transformation Technology)
Advisor	Assistant Professor Surapong Uttama, Ph. D.
Co-Advisor	Patcharaporn Panwong, Ph. D.

ABSTRACT

This research investigates the integration of cosmetic ingredient profiles and skin-related features to improve product recommendations utilizing machine learning and deep learning approaches. The core objective is to create comprehensive recommendations that tailor skincare product suggestions based on ingredient-focused machine learning algorithms, while lipstick recommendations with skin undertone and user preferences are made by utilizing deep learning techniques. The dataset for this research comprises three main components: Skincare product data, collected from a cosmetic website, this dataset includes detailed ingredient compositions, product features (product type, skincare ingredients). These characteristics serve as the foundation for content-based skincare recommendation. Skin Undertone Data, include wrist vein images, obtained from a public dataset, were used to classify skin undertones (cool, warm, neutral). These images were labeled with an expert and processed for deep learning input. Lipstick product data, collected from the e-commerce platform, contains lipstick finishes (matte, glossy, satin), benefit, ethical preferences and price. These features are critical for matching user preferences and harmonizing selections with skin undertone classification.

The research employs a two-part recommendation in cosmetic categories: Skincare recommendation with content-based filtering, and lipstick recommendation with deep learning and machine learning algorithms. In skincare product recommendation, ingredient compositions were used to recommend suitable skincare products for the user based on content-based filtering. In lipstick product recommendation, a convolutional neural network (CNN), Mobile Net V2, and

DenseNet121 architectures were trained to classify wrist vein images into skin undertone categories. Once undertone was identified, lipstick preferences (such as finish or benefit) were matched using clustering and content-based recommendation that maps product features to undertone-compatible options. For skincare recommendation, content-based filtering using cosine similarity and Jaccard similarity applied to structured ingredient data achieved accuracy of 80% in two skincare categories. The deep learning model for undertone classification achieved an accuracy of 84%, showing superior performance in handling vein color subtleties. In the lipstick recommendation, user preferences and product attributes were mapped against classified skin undertones with accuracy of 83% with content-based filtering.

Collectively, this research demonstrates the effectiveness of combining ingredient-based analysis for skincare products and image-based skin tone classification for lipstick products to deliver precise and personalized cosmetic product recommendations. By leveraging multimodal data sources and tailored algorithmic strategies, this research bridges technical rigor with cosmetic industry. The proposed framework offers valuable potential for intelligent beauty advisory recommendation approach, promoting data-driven decision-making in skincare and makeup personalization.

Keywords: Cosmetic Product Recommendation, Content-based Filtering, Machine Learning, Deep Learning

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND SYMBOLS

CF	Collaborative Filtering
CBF	Content-Based Filtering
TF-IDF	Term Frequency-Inverse Document Frequency
CNN	Convolutional Neural Network
RNNs	Recurrent Neural Networks
LSTMs	Long Short-Term Memory Networks
GANs	Generative Adversarial Networks
ReLU	Rectified Linear Unit
DBSCAN	Density-Based Spatial Clustering of Application with Noise
WOA	Whale Optimization Algorithm
WCSS	Within-Cluster Sum of Squares
ε	Epsilon or eps
MinPts	Minimum Number of Points
NLP	Natural Language Processing

CHAPTER 1

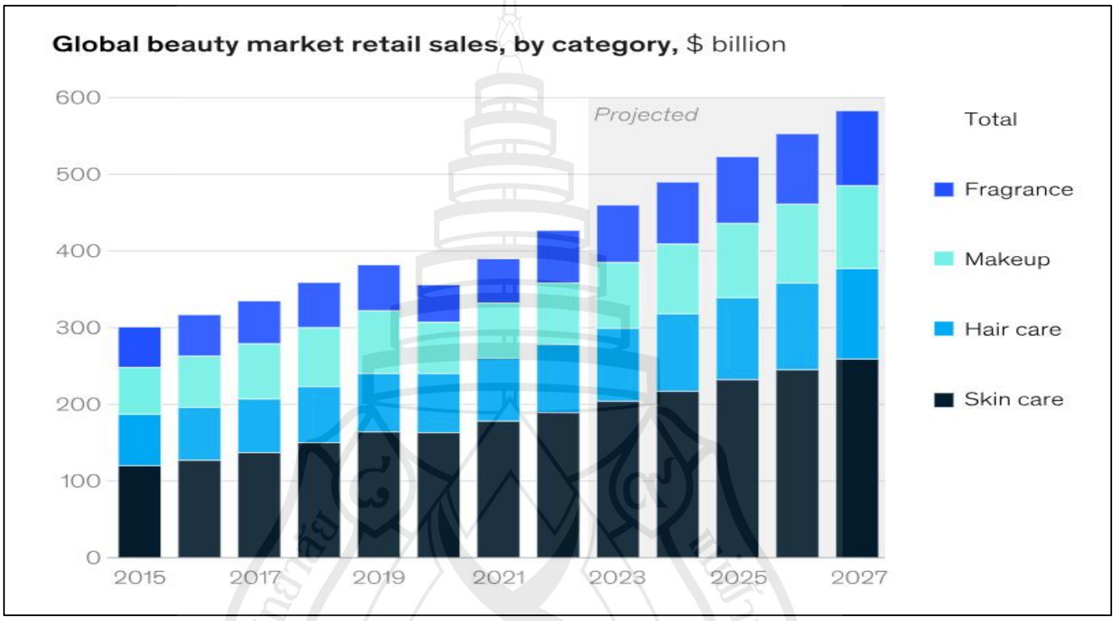
INTRODUCTION

1.1 Background Knowledge of the Research Problem

Cosmetic products are substances or mixtures applied to the human body, particularly the skin, hair, nails, lips, or teeth, for cleansing, beautifying, altering appearance, or enhancing attractiveness. Cosmetic products serve a variety of purposes, from enhancing physical appearance to maintaining skin health. They include a large range of items such as skincare creams, makeup, perfumes, deodorants, and hair dyes.

Cosmetics have been used for thousands of years, with evidence of early formulations discovered in ancient civilizations (Britannica, n.d.). In 4000 BCE, Egyptians were among the first to wear cosmetics, employing natural pigments as well as substances for beautification and ceremonial purposes. In Mesopotamia, women manufactured some of the earliest recorded cosmetics around 3200 BCE. In Greece, cosmetics were used for personal beautification and burial rituals, with recipes often including exotic ingredients. Roman women adopted Greek cosmetic practices, using makeup to signify wealth and status. After the fall of the Roman Empire, the Catholic Church discouraged cosmetic use, labeling it as sinful. However, this did not entirely stop their use, especially among lower classes and specific professions. In the 14th–17th century, cosmetics regained popularity in Europe, with aristocratic women using powders and rouges to create pale complexions. In the modern era (19th century onward), the industrial revolution brought significant advancements in cosmetic formulations (Cartwright, n.d.).

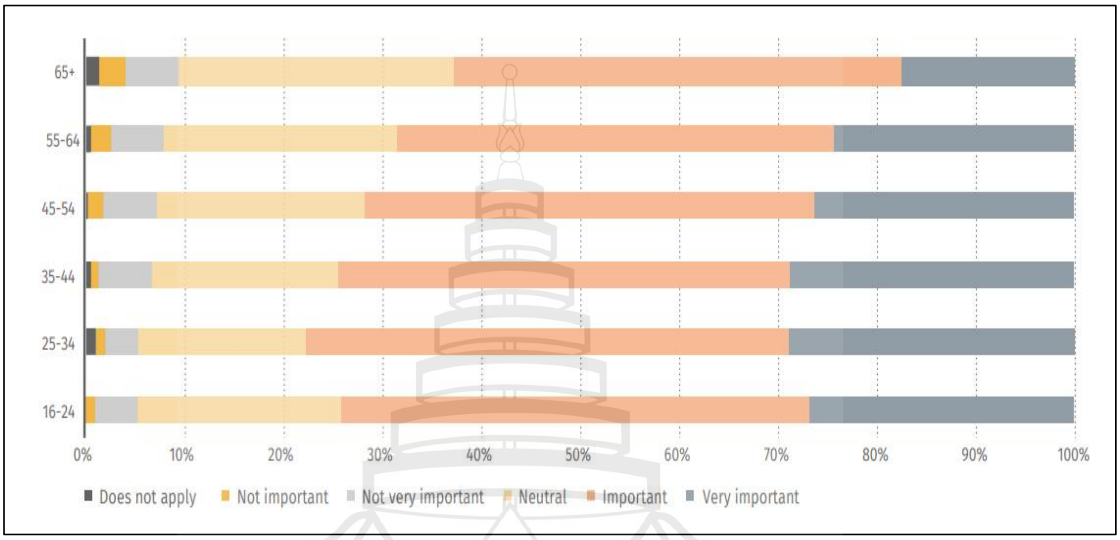
The cosmetic industry plays an important role in modern society, blending science, art, and commerce to cater to the diverse beauty and personal care needs of consumers worldwide. Encompassing a large range of products, together with skincare, makeup, haircare, as well as fragrances, the industry has become a dynamic and rapidly growing sector, driven by evolving consumer preferences, technological advancements, and increasing global awareness of self-care and personal grooming. The importance of the cosmetics industry extends beyond its economic contribution, which is substantial. Figure 1.1 represents the global cosmetic market by category from 2015 to 2027. Based on this information, the global cosmetics market was valued at USD 430 billion in 2022 and is predicted to grow at a rate of 6% per year to reach USD 580 billion by 2027.



Source Berg et al. (2023)

Figure 1.1 Global Cosmetic Market by Category

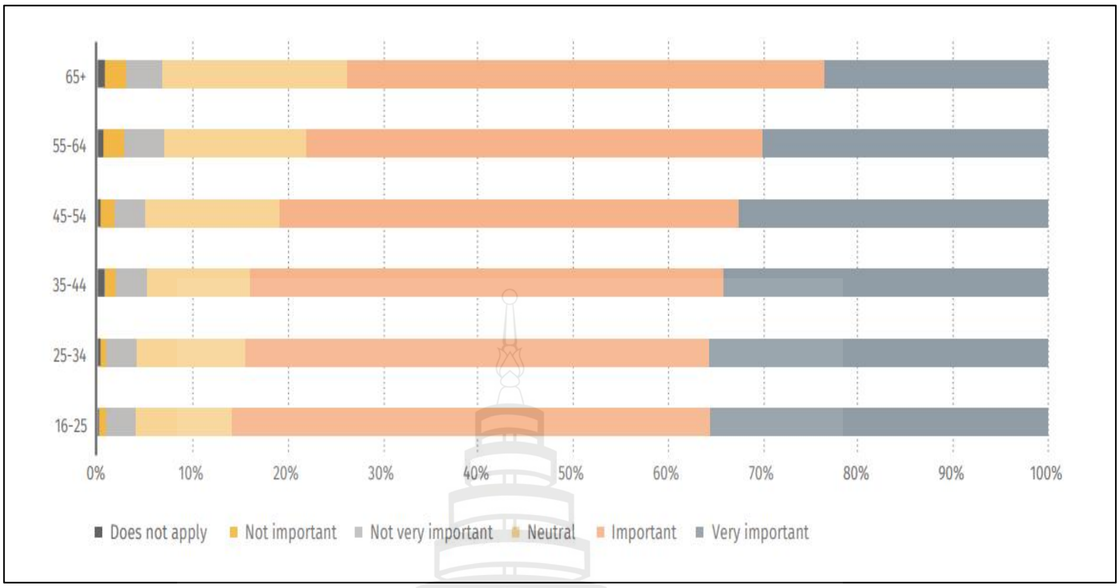
One of the key aspects driving market expansion is consumer awareness of ways to improve their personal appearance. Skincare, makeup, and hair care products are increasing in popularity as essential components of millennial grooming routines (Grand View Research, 2024). From a societal perspective, the cosmetic industry assists in improving emotional health, confidence, and self-worth. Skincare and makeup products are not merely tools for improving physical appearance but are also instruments for expressing individuality and cultural identity. Skincare products, for instance, address concerns such as acne, dryness, or aging, helping individuals maintain healthy skin and boost their psychological well-being. Makeup products can express and enhance the consumers' appearance, which can also influence personal and professional interactions. Consequently, cosmetic products hold both functional and emotional importance in consumers' lives. Moreover, the growing awareness of the link between health and beauty has propelled a demand for products that incorporate safe, sustainable, and dermatologically effective ingredients (Korichi et al., 2009).



Source Consumer Insights (2017)

Figure 1.2 Importance of Personal Care and Cosmetic Products by Age Group

According to consumer insights from Cosmetics Europe, 72% of individuals believe that beauty and personal care enhance their overall standards. By age group, Figure 1.2 shows the relative importance of cosmetics and personal care items in improving quality of life. Also, customers of all ages regard products as significant and useful for improving their relationships with others and boosting their sense of self. Figure 1.3 presents the study on the role of cosmetics and personal care items in boosting self-esteem by age group (Consumer Insights, 2017).



Source Consumer Insights (2017)

Figure 1.3 The Value of Cosmetics and Personal Care Items in Increasing Self-esteem on Age Group

Despite the growth and innovation, the cosmetic industry raises challenges to consumers. The market's vast diversity, with thousands of brands and goods, often overwhelms consumers, making it challenging to locate products that specially fulfill their specific needs. For example, consumers with sensitive skin must decipher ingredient labels to avoid severe reactions, whereas cosmetic users struggle to find products that match their complexion or undertones. The abundance of options complicates the selection process for consumers, who frequently struggle to find the solutions that are matched to their needs and preferences. Furthermore, marketing transparent information about product ingredients can confuse and mislead consumers. A survey by McKinsey & Company found that 70% of consumers are overwhelmed by product choices and find it challenging to select the right cosmetic products (Sexton, 2022).

Source Consumer Insights (2017)

Figure 1.4 Various Cosmetic Products

As a result, recommendations have evolved as helpful techniques for improving consumer experience by proposing products that meet their needs. Thus, this study will focus on two primary areas of the cosmetic industry: skincare and lipstick categories, to assist cosmetic consumers not only for health but also for their appearance by bridging the gap between their preferences and effective product selection.

1.2 Research Objectives

The study aims to create a recommendation methodology for cosmetic products that are suitable for consumers. The study's aims are as follows:

1.2.1 To collect and create the skincare and lipstick product dataset from the e-commerce website and skin undertone dataset.

1.2.2 To propose a content-based recommendation methodology for skincare products with their ingredient profiles.

1.2.3 To propose a recommendation methodology for lipstick products according to user preferences and skin-related features such as skin undertone.

1.2.4 To assess the effectiveness of proposed methods for cosmetic product recommendation.

1.3 The Importance of Research

The growing demand for cosmetic solutions underscores the importance of developing intelligent recommendations that align with consumer preferences and skin-related features. In skincare categories, product effectiveness heavily depends on the compatibility of ingredients. By leveraging ingredient profiles, this research proposes a recommendation that can scientifically recommend skincare products based on consumers' specific dermatological needs. This approach helps users steer clear of potentially irritating or useless ones while also boosting the accuracy of product recommendations.

In the makeup categories, particularly for lipstick recommendations, selecting the right shade based on skin undertone is a challenge many consumers face. Skin undertones are subtle yet essential in determining the most flattering shades, and finding a matching lipstick is often a trial-and error process. This research leverages deep learning techniques to automate and improve the process of classifying undertones, providing more precise lipstick suggestions. By doing so, it not only improves user satisfaction but also reduces the frustration of purchasing products that don't match or enhance a person's skin undertone.

By integrating content-based filtering and deep learning, the study demonstrates the potential for more accurate, scalable, and efficient methods of product recommendation. Additionally, the significance of this study extends beyond consumer satisfaction to the broader cosmetics industry by utilizing machine learning as well as deep learning to create new standards for beauty product recommendations.

1.4 Research Questions

The following are the research questions that the study seeks to answer:

1.4.1 How can ingredient profiles of skincare products be effectively utilized to recommend products?

1.4.2 How can the identified skin-related features (skin undertones) be integrated into a recommendation to suggest lipstick shades?

1.4.3 How can machine learning and deep learning techniques be enhanced by cosmetic product recommendations?

1.5 Scope of Research

This study emphasizes presenting a recommendation methodology for two primary areas using machine learning and deep learning approaches. These two primary areas are ingredient-based skincare recommendations and user preferences and skin-related features for lipstick recommendations. The study collected skincare products information from an e-commerce website for content-based skincare product recommendations, as well as skin undertone data from a public dataset and lipstick information from e-commerce websites to use deep learning to classify skin undertones and choose lipstick shades.

As for content-based skincare product recommendations, the skincare product dataset consists of 45 items with 5 features that were collected from an e-commerce website. For lipstick recommendations, the dataset contains 107 wrist vein images for skin undertone classes (warm, cool, and neutral). The wrist vein images were collected from a public dataset. The lipstick product dataset includes 117 items with 9 features that were collected from e-commerce. To support and validate the recommendation approaches, a survey conducted with 30 participants, primarily women aged 18-35, who are active users of cosmetic products. Their responses were used to validate the recommendation approaches, ensuring its alignment with user preferences and offering meaningful advancements in AI-driven technology in the beauty industry.

1.6 Research Limitations

A notable limitation in leveraging cosmetic ingredient profiles and skin-related features for content-based cosmetic product recommendations is the scarcity of prior research in this domain, which restricts the availability of established methodologies and benchmarks. Besides, there are not plenty of publicly available datasets containing detailed skin undertone labels, making it challenging to train and evaluate recommendation models effectively. This absence necessitates the creation of custom datasets, which is time-consuming and resource intensive. Reliance on a relatively small data set slightly affects the robustness of the trained models, as they may not effectively capture complex patterns.

1.7 Expected Outcomes

The expected outcomes of leveraging cosmetic ingredient profiles and skin-related features for cosmetic product recommendation are twofold. First, for skincare products, by analyzing ingredient profiles with content-based filtering, users can receive product suggestions that optimize skin health. Second, for lipstick recommendations, using deep learning for matching shades to users' skin undertones and considering user preferences to support the better recommendation results. This dual approach is expected to provide a more holistic and precise recommendation system, enhancing the overall user experience by offering products that align with both aesthetic preferences and individuals’ skin undertone.

CHAPTER 2

LITERATURE REVIEW

2.1 Theoretical Background

This study's theoretical framework integrates ideas from deep learning, content-based filtering, and recommendations for cosmetic products.

2.1.1 Product Recommendations

Recommendations have gained widespread popularity and are now used across many different fields. Recommendation methods are essentially methods for filtering and retrieving information and aim to provide users with relevant suggestions. These suggestions help guide users in making decisions, such as choosing which news article to read, what products to buy, or which songs to listen to. These recommendations can be personalized, offering different suggestions to different users, or non-personalized, providing the same suggestions to everyone. Recommendation methods are typically categorized based on how they generate suggestions, with the most common approaches being collaborative filtering (CF), content-based filtering (CBF), as well as hybrid filtering methods (Shah et al., 2017).

A well-liked recommendation method called collaborative filtering makes predictions about a user's preferences through analyzing the preferences and actions of comparable persons or objects. It works under the presumption that individuals who have interacted similarly in the past will likewise have similar preferences in the future. CF is effective in domains where user-item interactions (e.g., ratings or clicks) are plentiful. However, it has two problems: the sparsity issue (limited data available for accurate recommendations) as well as the cold-start issue (challenges recommending new users or products). To produce recommendations, content-based filtering uses user profiles and item attributes. This approach performs well in domains with rich metadata and does not depend on users' data, steer clear of cold-start problems for users (Mouhiha et al., 2024). To overcome the drawbacks of both CF and CBF approaches, hybrid systems combine them. These approaches enhance recommendation accuracy by exploiting collaborative relationships between users and items while also using item-specific attributes. The hybrid methods employ various strategies, such as blending predictions, switching between methods based on context, or integrating results. These approaches have been widely used across domains like movies, e-commerce, and more recently, cosmetics and personalized recommendations (Fahad Iqbal & Gnanajeyaraman, 2023).

2.1.2 Content-Based Filtering for Product Recommendation

By analyzing data features or items' content in conjunction with the user's previous choices, content-based filtering generates recommendations. In content-based filtering, item features as well as user preferences are utilized to recommend similar items. The main concept is to recommend items based on items that the consumer has already liked. Far from collaborative filtering, that relies on the multiple users' behavior, content-based filtering emphasis solely on the connection with a user's profile and item characteristics. This method avoids the cold start problem that can occur with newly added products that have not yet received user ratings, and it works particularly well in cases when user data is minimal. Content-based filtering normally consists of two primary steps: feature extraction, which mostly works TF-IDF, as well as calculation of similarity using similarity metrics (Lops et al., 2011).

A statistical measure known as Term Frequency-Inverse Document Frequency, or TF-IDF, is utilized to evaluate the value of a word within a document in relation to a set of documents. It is a key aspect of content-based filtering, especially in recommendations that use textual data to find similarities between items like product descriptions, reviews, or ingredient lists. TF-IDF converts textual data into numerical vectors. These vectors allow for item comparison and similarity evaluation by indicating the relative relevance of certain terms in the documents. In depth, the term frequency (TF) calculates the quantity of times a phrase appears throughout a document. Usually, it is normalized to avoid partializing lengthier documents. The inverse document frequency (IDF) calculates the significance of a term across the entire corpus. Terms like "and" and 'the" that appear frequently throughout every document are assigned a lower weight, while terms that are rarer are assigned a higher weight. When TF and IDF are multiplied, a score indicates how important a term is to the entire corpus in a specific document (Scott, 2021). The TF-IDF values are calculated using (1), (2), and (3) (Kumar et al., 2020).

$$TF(t,d) = \frac{Number\ of\ times\ t\ appears\ in\ d}{Total\ number\ of\ terms\ in\ d} \tag{1}$$

$$IDF\ (t) = log\ (\frac{N}{1 + (df \times t)}) \tag{2}$$

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \tag{3}$$

Where:

d = a document

N = the quality of documents

df = the total quality of documents with term t

Similarity metrics are used to determine the similarities between vectors based on TF-IDF. Cosine similarity is a metric that determines how similar two vectors are by computing the cosine of the angle between them. The formula of cosine similarity is shown in (4) (Abdurrafi & Ningsih, 2023).

$$Cosine\ Similarity(A,B) = \frac{A \cdot B}{\parallel A \parallel \ \ \parallel B \parallel} \tag{4}$$

Where:

$A \cdot B$ = the dot product of vector A and B

$\parallel A \parallel \ \ \parallel B \parallel$ = the magnitudes of vector A and B

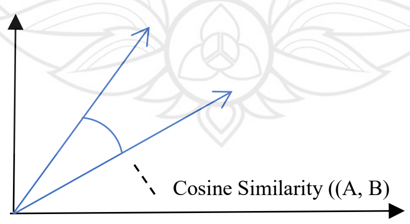


Figure 2.1 Cosine Similarity between Vector A and B

The value of the cosine of an angle formed by two vectors ranges from -1 to 1. A cosine similarity of one indicates that the vectors are similar. Meanwhile, 0 denotes orthogonality (no resemblance), and -1 represents completely dissimilarity. Cosine similarity is applied in content-based filtering to compare similarities of items, which are commonly represented as TF-IDF vectors or other feature vectors. Items with higher cosine similarity scores are considered more similar, making them ideal candidates for recommendation (Liu et al., 2024).

Another measurement called Jaccard similarity is utilized to access the similarity in two sets. It is often utilized in content-based filtering, especially when comparing how similar the two sets are: like user preferences and item features. The similarity is calculated as the ratio of shared features to the total number of features in the comparison sets. A score of 0 indicates no similarity, and a score of 1 indicates total similarity. More similarity between the products would be indicated by a higher Jaccard similarity score. The mathematical term of the Jaccard similarity is shown in (5) (Ziogas et al., 2022).

$$J(A, B) = \frac{/ A \cap B /}{/ A \cup B /} \tag{5}$$

Where:

$J(A, B)$ = Jaccard similarity of sets A, B

$| A \cap B |$ = the quantity of elements in the intersection of the two sets

$| A \cup B |$ = the quantity of elements in the union of the two sets

In some content-based recommendations, Jaccard similarity was used in conjunction with TF-IDF to enhance a recommendation. While TF-IDF captures the importance of terms across the dataset, Jaccard similarity applied to the binary version of these vectors to compare the presence or absence of ingredients between products (Jain et al., 2017).

2.1.3 Skin Undertone Classification for Cosmetic Matching

A fundamental principle in cosmetic artistry and formulation is the harmonious interplay between product shades and an individual's natural skin undertone. Skin undertone is a foundational aspect of human complexion, defined as the subtle, underlying tone beneath skin's surface. It remains consistent regardless of changes in skin color due to tanning or other external factors. Understanding skin undertones is essential in cosmetics, particularly in choosing complementary lipstick shades, as it directly influences how colors appear on the skin and enhances facial aesthetic.

Skin undertones are generally categorized into three types: warm, cool, and neutral. Warm undertones are characterized by golden, peachy, or yellow hues and are complemented by warm lipstick shades such as corals, oranges, and brick reds. Cool undertones, identified by pink, red, or bluish hues, pair well with shades like berry, rose, and blue-based reds. Neutral undertones, which exhibit a balance of warm and cool characteristics, have versatility and can adapt to a broader range of lipstick shades (Sirisayan, 2022).

The compatibility between skin undertone and lipstick shade is rooted in the principles of color theory. According to Johannes Itten's color wheel, analogous and complementary color schemes are key to creating visually harmonious combinations. When the lipstick shade aligns with the skin's undertone, it enhances the wearer's natural complexion, creating a balanced and vibrant look (Itten et al., 1996). Conversely, mismatched undertones can lead to a washed-out or overly stark appearance.

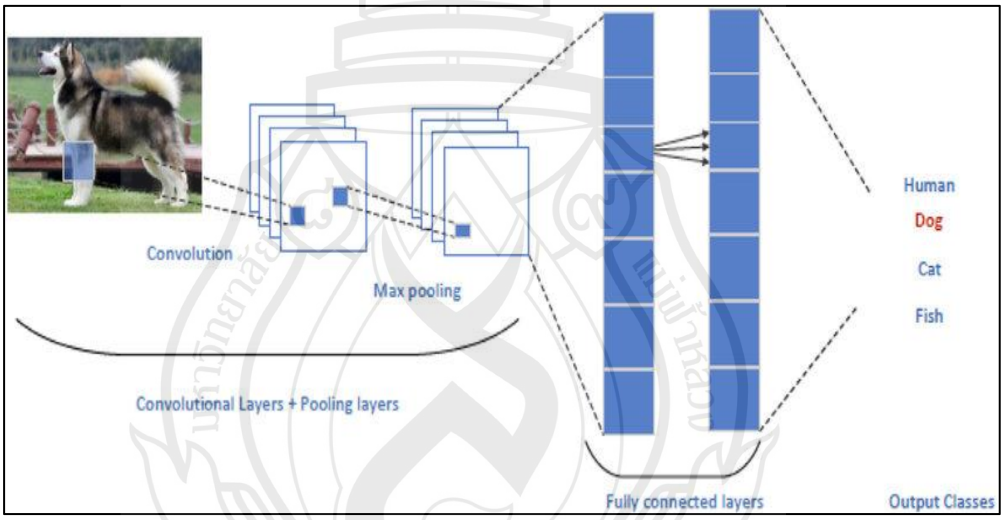
Understanding the relationship between skin undertone and lipstick shade is not only important for enhancing individual appearance but also for advancing product design in the beauty sector. As consumer preferences increasingly lean toward tailored solutions, recognizing the nuances of undertone compatibility continues to play a pivotal role in cosmetic innovation.

2.1.4 Deep Learning for Image Classification

Known as artificial neural networks, deep learning, a specific subclass of machine learning which draws inspiration from the structure and operation of the human brain. It works by building several interconnected node layers, each of which takes the raw input data and extracts higher-level features. This hierarchical learning method enables deep learning models to perform well in a variety of domains, including image identification, natural language processing, and recommendation systems. Deep learning algorithms are diverse based on the use of specific tasks of domains and types of data. CNNs, RNNs, LSTMs, GANs, autoencoders, and transformer networks are examples of deep learning algorithms that are fundamental for numerous applications (Manakitsa et al., 2024). This study focused on three models Traditional CNN, Mobile Net, and Dense Net are described below.

2.1.4.1 Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs), a popular deep learning model, work especially well for tasks involving images, such recommendation and classification. CNNs are intended to simulate visual processing in human brains. A CNN is composed of several layers, including the input layer, the output layer, the hidden layer (which usually includes convolution layers, pooling layers, and fully connected layers) and the hidden layer. The architecture of image classification using CNN is shown in Figure 2.2.



Source Laamouri and Nawal (2024)

Figure 2.2 CNN Model

Data is received by the input layer. The formula for the convolution layer, which is in the state of feature extraction, is provided in equation 6.

$$(D * C)(i,j) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} D(i+x, j+y) * C(x,y) \tag{6}$$

Where:

D = the data input

C = the filter or kernel

i and j = the convolution result's coordinates between D and C

N and M = the image width and height dimension

x and y = the matrix iteration indices

The pooling layer's formula is given in equation 7, and its purpose is to down sample features by reducing their size to speed up and improve the efficiency of the previous layer's activities.

$$(Pool_{max})(i,j) = max_{x=0}^{N-1} max_{y=0}^{M-1} D(i+x, j+y) \tag{7}$$

Where:

I and j = the pixels positions that come from the max pooling process

D = the data input

N and M = the pooling dimension (2x2 or 3x3 are often used sizes)

Based on the results of the CNN's pooling and convolution layers, the classification process is carried out using the fully connected layer. In equation 8, the fully linked layer formula is stated.

$$X_y = \sum_{x=1}^{N} (W_{x,y} * Actv_x) + B_y \tag{8}$$

Where:

N = the number of neurons in the layer above.

W(x,y) = the weight that connects a neuron in the layer that comes before it (x) to the layer that comes after it (y).

Actv(x) = the activation function in the preceding layer (x).

B(y) = the activation function

2.1.4.2 Mobile Net V2

Mobile Net is a family of lightweight convolutional neural network (CNN) architectures developed by Google, primarily developed for effective image classification on mobile and embedded systems. The original MobileNetV1 incorporated depth wise separable convolutions, which significantly reduced computing costs compared to traditional CNNs. Subsequent versions, such as MobileNetV2 and V3, incorporated innovations like inverted residuals and squeezeand-excitation modules to enhance performance and efficiency. These architectures have been widely adopted in applications requiring real-time image classification with limited computational resources (Chen et al., 2022).

The fundamental concept of MobileNetV2 is the use of depth-wise separable convolutions, which split the traditional convolution operation into distinct depth-wise and point-wise convolutions. By breaking down the procedure, MobileNetV2 significantly lowers the number of parameters and computing complexity while maintaining a notable level of accuracy. Depth-wise convolutions perform spatial filtering independently for each input channel, whereas point-wise convolutions aggregate the filtered outputs across channels.

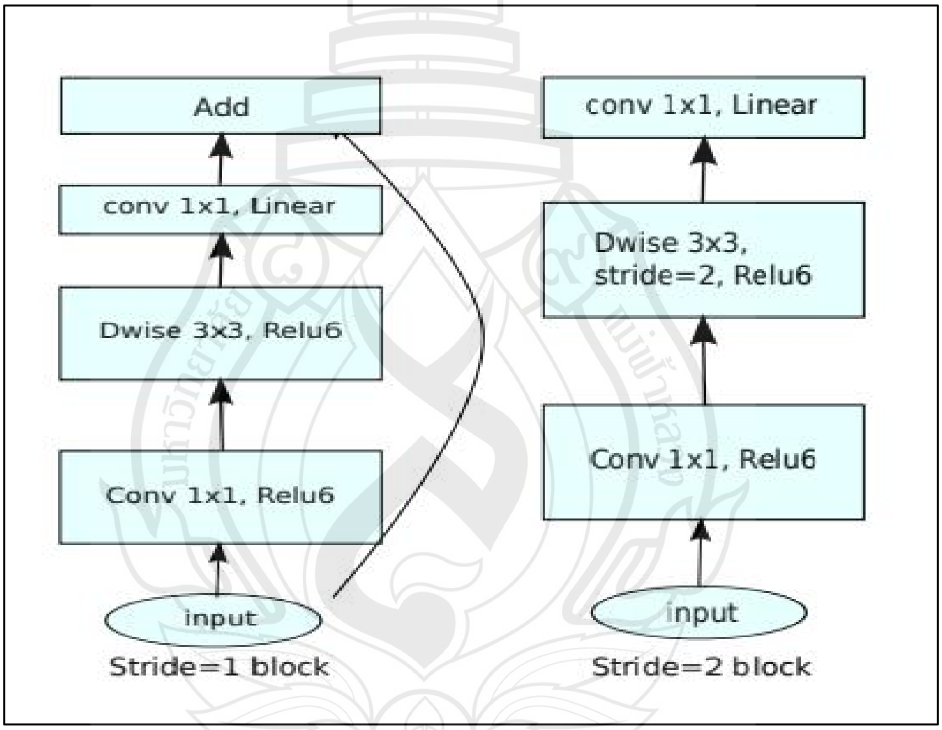
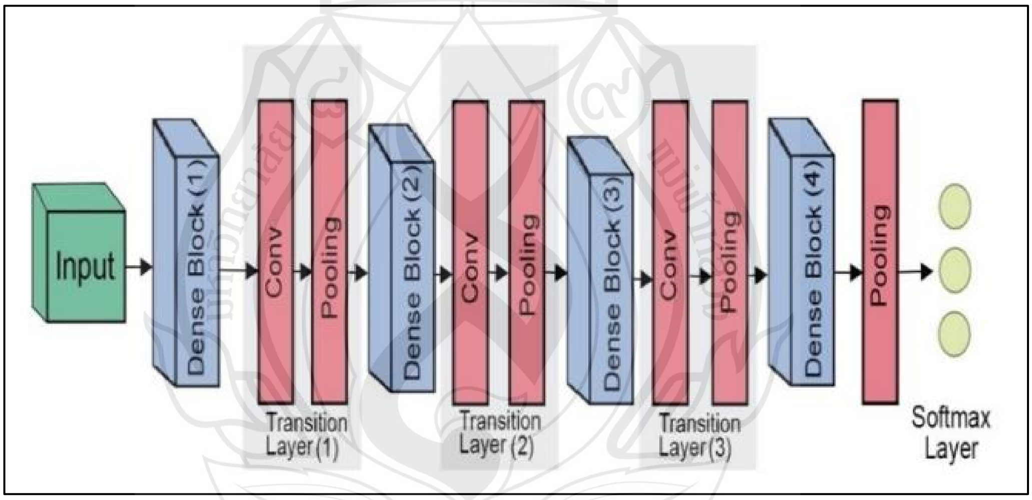


Figure 2.3 Mobile Net Architecture

The addition of inverting residual blocks, which lowers the network's computational expenses while boosting its representational capacity, is a significant improvement over MobileNetV2. Inverted residuals compress input feature maps using 1x1 convolutions using a bottleneck structure prior to performing a depth-wise convolution. Next, 1x1 convolutions are used to expand the dimensions. This bottleneck design facilitates effective information flow throughout the network while lowering computation. The architecture of Mobile Net V2 is demonstrated in Figure 2.3 (Sandler et al., 2019).

2.1.4.3 DenseNet121

An architecture called DenseNet121 is a CNN (Convolutional Neural Network) constructed to enhance feature utilization and facilitate information flow across network layers. In this architecture, each layer or block connects to every other layer or block through a feed-forward method. Each Dense Net layer includes three main elements: a convolution operation utilizing 3×3 filters, the ReLU activation function, as well as batch normalization. Batch normalization enables higher learning rates during model development and expedites the training process. A matrix of image pixels is provided as input to each block. Batch normalization is then used to process this matrix to lessen the likelihood of overfitting during the training phase. The architecture of DenseNet121 is shown in Figure 2.4 (Ariawan et al., 2025).



Source Albelwi (2022)

Figure 2.4 DenseNet121 Architecture

2.1.5 Clustering for Product Recommendations

A basic unsupervised machine learning method called clustering aims to arrange a collection of items so that those in the same group (cluster) are more alike than those in other groupings. Because clustering doesn't require pre-labeled data like supervised learning does, it's incredibly useful for anomaly identification, pattern recognition, and exploratory data analysis. Various algorithms exist, each with their own approach to defining similarity and forming clusters. Common types include density-based techniques like DBSCAN, which determines clusters based on the data points' density; model-based techniques, which presume that data points are produced through a varity of probability distributions; hierarchical techniques, which create a tree-like structure of clusters; and partitioning techniques like K-means, which divide data into a predetermined number of clusters. The type of data, the intended cluster shape, and the computational limitations all influence the clustering algorithm selection (Hu et al., 2024). This study concentrated on three models based on the dataset's nature: K Means Clustering, Density-Based Spatial Clustering of Applications with Noise and Whale Optimization Algorithm for K-Medoids Clustering, each of which is specified below.

2.1.5.1 K-Means Clustering

K-Means clustering, a popular unsupervised machine learning technique, separates n observations into k clusters, with each observation assigned to the cluster with the nearest mean (centroid). The primary goal of K-Means is to identify the centroids of each cluster by minimizing the total of squared distances of the assigned cluster centroid and the data points. The first step in this iterative approach is to initialize k centroids at random. In the "assignment stage," each data point is then allocated to a cluster that has the closest centroid to it. Subsequently, in the "update stage," the centroids are recalculated using the average of all that cluster's data points. These two procedures are carried out repeatedly until either a maximum number of iterations is reached, signifying convergence, or the cluster assignments stop changing. The K-Means algorithm intends to decrease inertia, or the within-cluster sum of squares (WCSS). The formula for WCSS is given by:

$$J = \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \sum_{\boldsymbol{x \in S_i}} ||x - \mu_i||^2 \tag{9}$$

Where:

J	= the objective function that needs to be reduced
k	= quantity of clusters
Si	= the collection of data points that make up the cluster i
x	= data point

μ_i = the center of cluster i

$x - \mu_i$ = the squared Euclidean distance between data point x and centroid $\boldsymbol{\mu_i}$ (Herdiana et al., 2025).

2.1.5.2 Density-Based Spatial Clustering of Application with Noise (DBSCAN)

A potent clustering algorithm called Density-Based Spatial Clustering of Applications with Noise (DBSCAN) organizes data points according to density rather than preset cluster numbers. In contrast to K-Means, DBSCAN can identify clusters of arbitrary shape as well as effectively handle noise by distinguishing outliers from dense regions. The algorithm works by defining two parameters: MinPts, the smallest number of points needed to construct a dense zone, and ε (epsilon), which determines the neighborhood radius. DBSCAN starts by selecting a random point and expanding a cluster if enough neighboring points exist within **ε**. If a point lacks sufficient neighbors, it is labeled as noise. This approach makes DBSCAN particularly suited for applications like anomaly detection, geographic data analysis, as well as image segmentation.

DBSCAN is mathematically based on the notions of density connectedness and density reachability. If a point q has at least MinPts neighbors and p is within ε of q, then p is directly density-reachable from q. An arrangement of points that are density-connected—that is connected by a chain of density-reachable points—forms a cluster. This density-based approach allows DBSCAN to discover complex structures in data without requiring prior knowledge of cluster numbers. The formal definition of DBSCAN clustering is expressed in equation 10 (Cheng et al., 2024).

$$N\,(p) = \{q \mid d(p,q) \leq \in\ \} \qquad (10)$$

Where:

N(p) = the neighborhood of point p, and d(p,q) is the Euclidean distance measure

2.1.5.3 Whale Optimization (WOA) K-Medoids

The metaheuristic optimization method known as the Whale Optimization Algorithm (WOA) was developed in response to humpback whale hunting behavior. The “bubble-net” feeding method, in which whales create spiral bubbles to catch prey, is replicated by this algorithm. In the context of K-Medoids clustering, WOA is employed to effectively search for optimal medoids, which are actual data points representing the centers of clusters. Unlike K-means, which uses centroids (means that may not be actual data points), K-Medoids is more robust to outliers as it selects existing data points as cluster representatives. The WOA-K-Medoids approach leverages the exploration and utilization capabilities to explore the search space of possible medoid configurations, with the aiming to minimize the dissimilarity between data points and their assigned medoids, thereby improving the quality of the resulting clusters. The mathematical equation for WOA is described in equation 11 (Chenan & Tsutsumida, 2025).

$$X\,(t+1) = X^{*} - A \cdot D \tag{11}$$

Where:

X* = the best solution found so far

A = a coffcient vector controlling exploration

D = the distance between the whale's current position and the best solution

2.1.6 Performance Evaluation Measures

Evaluation metrics are crucial for evaluating how well deep learning-based clustering algorithms and classification models work. Accuracy, Precision, Recall, F1-score, and Confusion Matrix are applied to assess the performance of classification models. The classification models' results can be categorized as either positive or negative. A positive categorization result indicates an accurate one, whereas a negative one indicates an inaccurate one. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are identified using precision, recall, and F1-score, a combination of positive and negative results in a more thorough prediction measure. The following assessment criteria were used in this work:

$$Precision = \frac{TP}{TP+FP} \tag{12}$$

$$Recall = \frac{TP}{TP+FN} \tag{13}$$

$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{14}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

Accuracy defined in equation (12) establishes the fraction of correctly classified instances in the overall dataset, which provides a general performance overview. Precision defined in equation (13) evaluates the fraction of correctly predicted positive instances among all predicted positives, making it useful in scenarios where false positives are costly. Recall defined in equation (14) assesses the fraction of correctly predicted positive instances among all actual positives, ensuring that the model captures as many relevant instances as possible. F1-score defined in equation (15) is the harmonic mean of precision and recall, balancing both metrics when dealing with imbalanced datasets (Terven et al., 2025).

The performance of the classification approach is summarized in a table called confusion matrix. It offers a thorough analysis of a set of occurrences' actual and predicted classes. The matrix is arranged in rows and columns, with the actual class represented by each column and the anticipated class by each row. The confusion matrix's entries indicate the number of instances, according to their actual and expected class labels, fit into categories. A prediction matrix assesses the accuracy of the model and aids in identifying the different kinds of errors. Comprehension of the strengths as well as limitations of the model is made possible by the patterns found in the matrix. A confusion matrix is a crucial tool for assessing performance in classification tasks since it offers a more comprehensive understanding of the model's performance. It serves as a foundation for computing additional evaluation measures, such as recall, accuracy, and precision. A confusion matrix provides a clear visual representation of the classification findings, allowing for a detailed analysis of the model's predictions and the identification of areas for improvement (Fahmy Amin, 2023).

For clustering algorithms, the silhouette index score is used to measure the condition of clusters. It calculates the degree to which each data point fits into its designated cluster in relation to other clusters. The Silhouette Score of a single data point is calculated by dividing its average distance from neighboring points (b) by its average distance from other points in the same cluster (a), and then dividing the result by the maximum of a and b. The formula for the point of data i is shown in equation 16.

$$S\,(i) = \frac{b(i)-a(i)}{max\,\{a(i),b(i)\}} \tag{16}$$

Where:

S = silhouette score

a = the average of intra-cluster distance

b = the average of nearest-cluster distance

i = data point

Higher scores on the Silhouette Score scale, which coordinates from -1 to 1, imply more distinct groupings. A score close to 1 suggests a good clustering since the data point is not well-matched to near clusters as well as well-matched to its possess cluster. Overlapping clusters are indicated by a silhouette score of about 0. If the silhouette score is close to -1, the data point may be assigned to the incorrect cluster. The average Silhouette scores of a clustering solution is the mean of the Silhouette scores on each of the points (Yin et al., 2024).

2.2 Related Works

This study utilized machine learning as well as deep learning techniques for effective cosmetic product recommendations in two primary areas: skincare and lipstick products. Several researchers worked on skincare product recommendations by focusing on item ingredients (Suvarna & Balakrishna, 2022) used NLP concept in content-based filtering for ingredients to recommend skincare products based on user skin type. They classified skin types into five groups and product categories into six categories, assigning each product's ingredients binary values. The top five items were suggested using cosine similarity method.

Similarity to this, (Lee et al., 2023) used the NLP concept for ingredients in content-based filtering to build skincare recommendations. Additionally, the researchers used the IF-IDF method to recommend products that have a user-desire beauty effect based on the product's ingredients. Another study (Iwabuchi et al., 2017) utilized IF-IDF filtering to suggest skincare with ingredients that have a high beauty effect. Their methodology takes into consideration reviews, the ingredients of skincare products, seven levels of satisfaction, and fifteen different types of effects (moisturizing, anti-aging, acne, etc.) The result showed that the reliability of the proposed methodology was unenforceable in each recommended product group of less than 5%.

Some researchers proposed content-based filtering using TF-IDF and cosine similarity in numerous recommendations (Lumintu, 2023) created a content-based personalized e-commerce recommendation utilizing TF-IDF as well as the cosine similarity method. The study processed product description data using TF-IDF and then computed the similarity product based on high cosine similarity score (Raihan et al., 2024) used TF-IDF and cosine similarity in designing hangout places recommendation applications to provide accurate and user-friendly recommendations. These two methods provided relevant recommendations and enhanced user satisfaction and engagement by aligning user preferences with hangout spot features (Permana & Wibowo, 2023) employed TF-IDF as well as the cosine similarity in content-based movie recommendations. The researchers focused on extracting keywords from movie synopses, integrating them with genres, and assessing the recommendation using performance metrics like F1-score and precision.

While it comes to cosmetic product recommendations, (Patty et al., 2018) applied content-based filtering for the purchasing cosmetic product recommendation. The study focused on the evaluation of item profiles and user profiles using TF-IDF and cosine similarity. However, they did not provide a numerical evaluation of the accuracy of their proposed method.

TF-IDF, cosine similarity, and Jaccard similarity were utilized for several recommendations (Tare et al., 2023) provided relevant event recommendations according to the user's location, interests, and history. Researchers calculated the word frequency of each event with TF-IDF and then compared it with another by using cosine and Jaccard similarities (Sukestiyarno et al., 2023) developed the content-based recommendation system for e-learning platform by applying cosine and

Jaccard similarities. The research discovered that cosine similarity yielded an average similarity value of 0.6, which was higher than Jaccard similarity, which was 0.3.

Personalized lipstick recommendations have gained traction in recent years, driven by advancements within artificial intelligence (AI) as well as deep learning. Tailoring lipstick choices to individuals based on skin undertone and user preferences ensures more accurate and satisfying recommendations. The ability to classify skin undertones accurately is fundamental to recommending suitable lipstick products. Several works have explored deep learning approaches for skin undertones classification (Fayyadhila et al., 2021) presented a deep learning approach utilizing Convolutional Neural Networks (CNNs) to classify skin undertones for makeup shade selection by using a wrist vein images dataset. The study included feature extraction using convolutional and pooling layers, data augmentation (rotation, zooming), and final classification using fully connected layers. However, the model was instructed on a relatively small dataset, including only 30 images per class before augmentation. This may restrict the ability of the model to generalize over diverse skin undertones and lighting conditions (Fortuna et al., 2024) focused on enhancing the accuracy of skin undertone classification and makeup recommendations by combining RGB and YCbCr color spaces with the convolutional neural network (CNN), addressing challenges related to lighting inconsistencies in previous study. The study employed wrist skin image dataset with visible veins (143 records). Metrics like accuracy, precision, recall, and F1-score were used to evaluate the model. As a result, it demonstrates that CNNs capture color complexities more effectively, leading to personalized and precise makeup suggestions. Nevertheless, the study does not integrate full recommendation and limiting datasets, compare CNN with other classification models, and integrate advanced image preprocessing techniques.

Existing research in skin undertone classification establishes a foundation for AI-driven cosmetics recommendations by accurately identifying individual undertones using deep learning techniques such as CNNs. Building on this, AI-driven personalization recommendation techniques refine the selection process, ensuring makeup choices align with user-specific characteristics by integrating image processing and preference-based algorithms (Cahyono et al., 2023) applied a Convolutional Neural Network (CNN)-based system for predicting lipstick color reproduction on users' lips. By analyzing RGB values of lips before and after applying lipstick, the model enhances virtual try-on experiences by providing more realistic color predictions. The research underscores the advantage of CNN over traditional methods in capturing spatial features that influence color application. However, limitations include data set constraints; it is relatively small, comprising only three lipstick colors and 35 participants, limiting the generalizability of the model.

Personalized cosmetic recommendations rely not only on image-based classification but also on user preference products. Therefore, clustering-based product recommendation plays a vital role in organizing and grouping cosmetic products, using methods like K-Means, DBSCAN and WOA K-Medoids to structure lipstick selections based on product features and user preferences (Khatreja & Mulay, 2022) applied an integrating K-means clustering for attribute categorization and semi-supervised learning. The system shows the effectiveness of combining machine learning techniques for cosmetic personalization, outperforming previous approaches that relied solely on RGB-based classification or K-means clustering. However, the study acknowledges limitations such as data scarcity and the need for real-world validation in recommendation effectiveness. The findings emphasize the practical applications of AI in the cosmetic industry, allowing users to receive personalized lipstick suggestions efficiently.

While existing studies have explored various facets of cosmetic product recommendation, this research addresses the gap within the current literature by adopting comprehensive and validated methodologies. For skincare product recommendations, the work augments current content-based filtering techniques by conducting a direct comparative analysis between cosine and Jaccard similarities for ingredient-based analysis, a crucial distinction often absent in prior skincare-focused investigations. Crucially, this is complemented by expert evaluation, thereby transcending the sole reliance on computational metrics. Regarding lipstick product recommendations, the research advances skin undertone classification beyond singular model approaches through the implementation and comparative assessment of Convolutional Neural Networks (CNN), Mobile Net, and Dense Net, promising more robust and generalizable models. Furthermore, this study overcomes the fragmentation of prior research by integrating deep learning for undertone classification with clustering and content-based filtering to incorporate user preferences, thereby establishing a more holistic recommendation framework. This integrated approach is further strengthened by the inclusion of expert validation, a vital component for ensuring real-world applicability that has frequently been understated in previous computational studies.

CHAPTER 3

METHODOLOGY

3.1 Overall Research Methodology

The study focused on the two primary areas of the cosmetic industry, skincare and lipstick categories, by utilizing machine learning as well as deep learning techniques. The study's proposed methodology is presented as Figure 3.1.

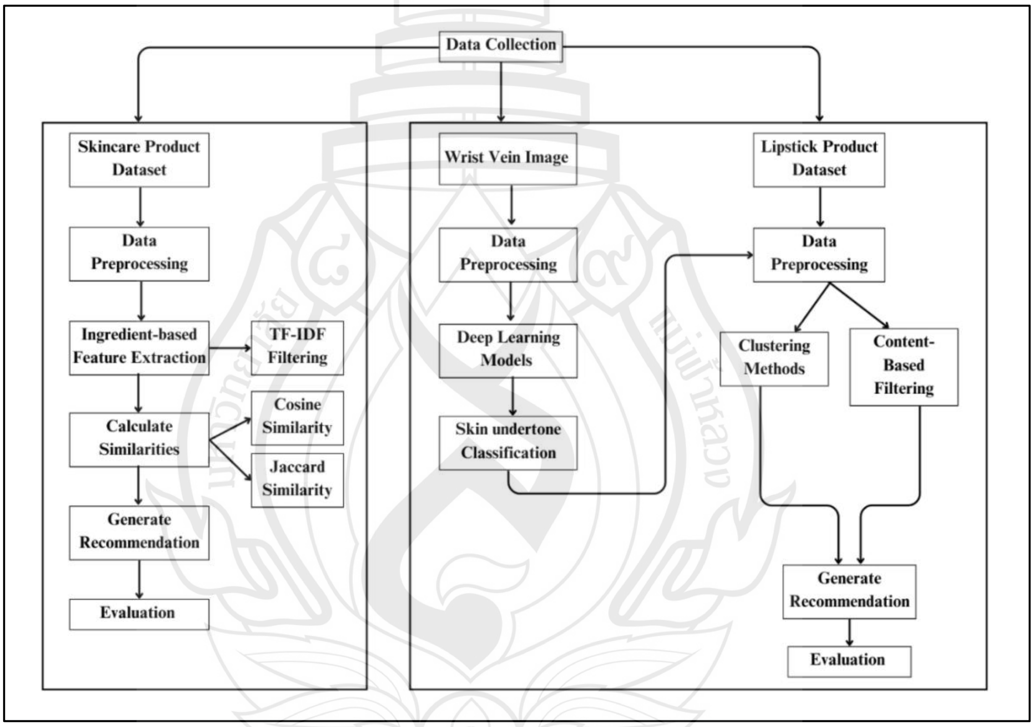


Figure 3.1 Research Methodology

The initial step is the gathering of skincare data from the e-commerce website that sells numerous types of cosmetic products. After this, extracting the product ingredients from skincare data is the second step. The third step is calculating the similarity between the products using similarity metrics. The following step is generating the top similar products between the products in the corpus. The final step is evaluating the performance of the recommendation method.

In lipstick recommendations, skin undertones were identified using the deep learning models by using wrist vein images, and suitable lipstick shades were recommended based on skin undertone and user preferences. Then, evaluate the performance of the model.

3.2 Research Methodology for Skincare Recommendations

3.2.1 Data Collection

The skincare data used in this study was collected from an e-commerce website called the Watsons cosmetic website. The dataset comprises 45 items in CSV format. Mainly, five categories, such as moisturizer, cleanser, toner, serum, and sunscreen. Each category includes 9 items with 'Categories', 'Name', Ingredients', Skin type', and 'URLs." After collecting the data, duplicate values were removed, and missing values were handled in the preprocessing step. The attributes of the skincare dataset are expressed as Table 3.1, and the example of the skincare dataset is shown in Figure 3.2.

Table 3.1 Features of the Skincare Dataset

Name	Description
Categories (text)	Product Type
Name (text)	Product Name
Ingredients (text)	Ingredients of the product
Skin type (text)	User Skin type for the user

Categories	Name	Ingredients	Skintype	Url
Moisturizer	Loreal Paris Glycolic-Bright Glowi	AQUA / WATER, GLYCERIN, NIA	Normal	https://www.watsons.co.th/en/l-oreal-
Moisturizer	Cetaphil Moisturizing Cream. Dry	AQUA, GLYCERIN, PETROLATU	Dry, Sensitive	https://www.watsons.co.th/en/cetaph
Moisturizer	Neutrogena Hydro Boost Hyaluro	WATER, GLYCERN, DIMETHICO	All	https://www.watsons.co.th/en/neutro
Moisturizer	Revlon Evivesse Skin Reschedul	Aqua Mineral Oil Butylene Glycol	All	https://www.watsons.co.th/en/revlon-
Moisturizer	Bioderma Sensibio Defensive	AQUA/WATER/EAU, GLYCERIN,	Sensitive	https://www.watsons.co.th/en/bioder
Moisturizer	Olay Regenerist Micro-Sculpting	WATER, CYCLOPENTASILOXAN	All	https://www.watsons.co.th/en/olay-re
Moisturizer	MOISTURE SURGE™ EXTENDE	Water/Aqua/Eau, Dimethicone, Bu	Normal, Dry, Co	https://www.central.co.th/en/clinique-
Moisturizer	KIEHL'S ULTRA FACIAL CREAM	Water, Glycerin, Cyclohexasiloxar	Normal, Dry, Co	https://www.central.co.th/en/kiehls-ul
Moisturizer	LANCÔME ABSOLUE SOFT CR	Aqua/Water, Glycerin, Hydrogena	Normal, Dry, Co	https://www.central.co.th/en/lancome
Cleanser	CETAPHIL GENTLE SKIN CLEA	purified water, cetyl alcohol, propy	All	https://www.watsons.co.th/en/cetaph
Cleanser	Loreal Paris Glycolic-Bright Glowi	BAHAN AQUA/WATER GLYCERI	All	https://www.watsons.co.th/en/l-oreal-
Cleanser	Neutrogena Deep Clean Acne Fo	Water,Glycerin, Sodium Cocoyl G	Oily, Sensitive	https://www.watsons.co.th/en/neutro
Cleanser	Cerave SA Smoothing Cleanser	Salicylic acid/ Ceramides/ Niacina	All	https://www.watsons.co.th/en/cerave
Cleanser	La Roche Posay Effaclar Deep C	AQUA/WATER. GLYCERIN. MYR	Oily, Sensitive	https://www.watsons.co.th/en/la-roch
Cleanser	Revlon New Complexion Foamin	Water Aqua Glycerin Myristic Acid	All	https://www.watsons.co.th/en/revlon-
Cleanser	Simple Kind To Skin Refreshing F	Aqua, Cocamidopropyl Betaine, P	All	https://www.watsons.co.th/en/simple-
Cleanser	Dr.G pH Cleansing Gel Foam	Water(Aqua/Eau), Glycerin, Disod	All	https://www.watsons.co.th/en/dr.g-ph
Cleanser	Nivea White Oil Clear 5in1 Moistu	Aqua, Potassium Myristate, Propy	All	https://www.watsons.co.th/en/nivea-w
Toner	Neutrogena Ultra Gentle Alcohol-	Water, Glycerin, Glycereth-26, Po	Normal	https://www.watsons.co.th/en/neutro
Toner	Loreal Paris Glycolic-Bright Glowi	AQUA / WATER, GLYCOLIC ACI	Sensitive	https://www.watsons.co.th/en/l-oreal-
Toner	Cetaphil Bright Healthy Radiance	AQUA, BUTYLENE GLYCOL, NIA	All	https://www.watsons.co.th/en/cetaph
Toner	Manyo Factory Bifida Biome Amp	Water, Butylene Glycol, Glycerin,	All	https://www.watsons.co.th/en/manyo
Toner	Simple Soothing Facial Toner	Aqua, Hydrogenated Starch Hydr	All	https://www.watsons.co.th/en/simple-
Toner	Revlon Evivesse Milky Toner	Water Aqua Denatured Alcohol Pr	All	https://www.watsons.co.th/en/revlon-

Figure 3.2 Example of Skincare Dataset

Additionally, skincare specialist survey data for skincare selection was collected with an online survey method and stored in CSV format. The survey data was gathered from five skincare specialists who have high skincare knowledge. Especially, this dataset was acquired to assess the performance of the content-based filtering method.

3.2.2 Ingredients Feature Extraction

The ingredient extraction was executed with the Python programming language along with necessary libraries like NumPy, Pandas, Matplotlib and sklearn. Initially, all ingredients were extracted from the ingredient's column of the skincare dataset. The TF-IDF Vectorizer function from the sklearn library was then utilized in the recommendation to generate a representation of each ingredient category's important features. The function of TF-IDF Vectorizer is converting the collection of raw documents into a TF-IDF features matrix. This vectorizer computed the IDF (inverse document frequency) on the ingredient data of the skincare products. In this step, the vocabulary and the IDF values that were utilized to transform the text to a TF-IDF matrix were learned. After this step, map the features names: the ingredient terms that correspond to the columns for the TF-IDF matrix. The next step was transforming the ingredient text data into a matrix of TF-IDF features using the learned vocabulary and IDF values. The final step was showing the TF-IDF matrix. A sample of the TF-IDF matrix, where rows stand for products and columns for unique ingredients are mentioned in Table 3.2.

Table 3.2 TF-IDF Product-Ingredient Matrix

Product Name	Stearyl	Ingredients Disodium	Salicylate
Cetaphil Facial Gentle Skin Cleanser	0.2211	0.0000	0.0000
Simple Skin Refreshing Facial Wash	0.0000	0.1921	0.0665

3.2.3 Calculating the Similarities

The similarity within the products according to their ingredients lists was computed on the TF-IDF matrix utilizing the cosine similarity, which evaluates the cosine angle formed by two vectors pairwise. Next, a data frame was created from the cosine similarity matrix, with both rows and columns labeled by product names. The result for each product was computed as a similarity matrix, which indicates the similarity between two products. The similar scores close to 1 indicate that two products have similar ingredient profiles. The example of cosine similarity scores between pairs of products are described in Table 3.3.

Table 3.3 Cosine Similarity Product Matrix for Skincare Products

Product Name	Neutrogena Deep Clean Acne Cleanser	La Roche Posay Deep Cleaning Foaming	Loreal Paris Glycolic-Bright Cleanser
Cetaphil Facial Gentle Skin Cleanser	0.0755	0.0153	0.0558
Simple Skin Refreshing Facial Wash	0.2909	0.1330	0.2874

In addition to cosine similarity, Jaccard similarity is an alternative method used for computing the similarity of skincare products to another. To calculate Jaccard similarity, the TF-IDF matrix is converted into a binary matrix. Where each entry represents the presence (1) or lack (0) of an ingredient. Then, the similarity between two products was computed by using Jaccard similarity on the binarized TF-IDF matrix. The similarity between the two products was defined as the ratio of an intersection of ingredients to the union of ingredients. The Jaccard similarity between product pairs is illustrated as Table 3.4.

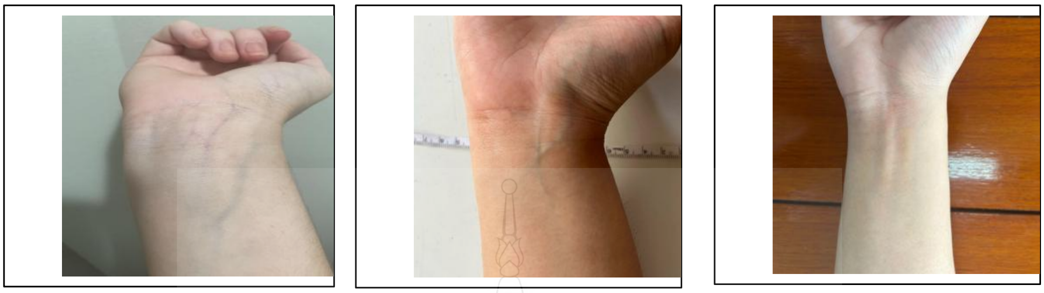
Table 3.4 Jaccard Similarity Product Matrix for Skincare Products

Product Name	**Neutrogena Deep Clean Acne Cleanser**	**La Roche Posay Deep Cleaning Foaming**	**Loreal Paris Glycolic-Bright Cleanser**
Cetaphil Facial Gentle Skin Cleanser	0.0714	0.0303	0.0526
Simple Skin Refreshing Facial Wash	0.1695	0.1282	0.0757

3.3 Research Methodology for Lipstick Recommendations

3.3.1 Data Collection

Following the skincare recommendation, data is gathered for the lipstick recommendation. The lipstick selection approach specifically uses the skin undertone criterion to recommend products that match the colors of each undertone category. Deep learning models were used to classify skin undertones, which were then incorporated into a content-based filtering process to suggest appropriate lipsticks based on product attributes and skin undertone. The skin undertone data based on the wrist vein image was collected from the Roboflow website. The dataset contains 107 wrist vein images in png format and labeled by a specialist. After labeling, the wrist vein images are divided into 3 skin undertone dataset, namely “Warm”, “Cool”, and “Neutral” include 70 images for warm skin undertone, 29images for cool skin undertone, and 10 images for neutral skin undertone. During the preparation stages, image augmentation and enhancement were performed to raise the quantity of the data to an appropriate level. The augmentation works with 90, 180, 270, 360-degree rotation with 700 images in each skin undertone group. The dataset was divided into three segments: 10% for validation, 10% for testing, and 80% for training. To classify skin undertones based on wrist vein images, a convolution neural network (CNN), Mobile Net V2, and DenseNet121 architecture were implemented with the Python programming language along with necessary libraries. Examples of skin undertone data are shown in Figure 3.3.



(a) Cool Skin Undertone (b) Warm Skin Undertone (c) Neutral Skin Undertone

Figure 3.3 Three Types of Skin Undertone

Additionally, lipstick product data were collected from the Sephora e-commerce website to support the lipstick recommendation system. The dataset includes 39 products for each skin undertone group. In the preprocessing stage, raw data were cleaned by removing duplicate entries and handling missing values to ensure efficient clustering and analysis. Key categorical features—Undertone, Finish, Benefit, and Ethical Preferences—were selected and transformed into numerical format using OneHotEncoder, which created separate binary columns for each unique category. Finally, the data were standardized to prepare the diverse product attributes for analysis, with the dataset features detailed in Table 3.5 and an example shown in Figure 3.4.

Table 3.5 Features of the Lipstick Dataset

Name	Description
Name (text)	Product Name
Brand Name (text)	Product Brand name
Color (text)	Color of the product
Undertone (text)	Suitable undertone of the product
Finish (text)	Finish of the product
Benefit (text)	Benefits of the product
Ethical Preferences (text)	Ethical preferences of the product
Price (integer)	Price of the product

Name	Brand Name	Color	Undertone	Finish	Benefit	Ethical Preferen	Ingredients	Price
SatinAllure™ Lip	PAT McGRATH I	Light Peachy Nu	Cool	Satin Finish	Hydrating	Vegan	Polybutene,	30$
Mineralist Hydra	bareMinerals	Pink Nude	Cool	Satin Finish	Hydrating, Long-	Vegan	Ricinus Cor	28$
Monochrome Hy	Prada Beauty	Mauve Nude	Cool	Matte Finish	Hydrating, Long-	Stick Formula	Dimethicone	50$
Long-Wearing M	Anastasia Bever	Pinky Brown	Cool	Matte Finish	Long-wearing	Cruelty-free, Alc	Octyldodec	23$
Prada Monochrc	Prada Beauty	Brown Nude	Cool	Matte Finish	Hydrating	Stick Formula	Dimethicone	50$
Kind Words Matt	Rare Beauty by	Berry Rose	Cool	Matte Finish	Hydrating, Long-	Vegan, Cruelty-fi	Isononyl Isc	20$
Satin Hydrating I	SEPHORA COL	Nude	Cool	Satin Finish	Hydrating, Long-	Vegan, Stick For	Polyglycery	16$
Powermatte Lon	NARS	Berry Pink	Cool	Matte Finish	Long-wearing	Stick Formula, W	Dimethicone	34$
Unlocked™ Sati	Hourglass	Warm Beige	Cool	Satin Finish	Full Coverage	Vegan, Cruelty-fi	Tridecyl Trir	38$
Major Headlines	PATRICK TA	Berry	Cool	Matte Finish	Hydrating, Long-	Vegan, Stick For	Cetearyl Eth	32$
Lip Suede Hydra	Westman Atelier	Red	Cool	Matte Finish	Hydrating	Vegan, Cruelty-fi	Isoamyl Lau	50$
Weightless Lip C	Kosas	Baby Pink	Cool	Satin Finish	Hydrating, Long-	Cruelty-free	Ricinus Cor	26$
Color Block High	ILIA	Bardot Nude	Cool	Satin Finish	Hydrating, Long-	Vegan, Gluten F	Ricinus Cor	28$
MAKEOUT CLU	Freck Beauty	Rose Brown	Cool	Matte Finish	Hydrating, Long-	Vegan, Cruelty-fi	Dimethicone	24$
Forget the Filler	LAWLESS	Soft Baby Pink	Cool	Satin Finish	Hydrating, Plump	Vegan, Cruelty-fi	Rincinus Cc	28$
Speak For Yours	Sarah Creal	Pink Nude	Cool	Satin Finish	Hydrating, Plump	Vegan, Cruelty-	Polybutene,	50$
Speak Love Moi:	LYS Beauty	Reddish Brown	Cool	Matte Finish	Hydrating, Long-	Vegan	Triisotridecy	20$
Unlocked™ Soft	Hourglass	Pink Beige	Cool	Matte Finish	Long-wearing	Vegan, Cruelty-fi	Dimethicone	38$
Satin Lip Color F	ROSE INC	Pink Beige	Cool	Satin Finish	Hydrating, Long-	Vegan, Stick For	Pentaerythr	16$
Matte Revolution	Charlotte Tilbury	Nude Peach	Cool	Matte Finish	Long-wearing	Vegan, Without I	Neopentyl C	35$

Figure 3.4 Example of Lipstick Dataset

Furthermore, user survey data was gathered using an online survey approach and stored in csv format. The survey responses were utilized to extract meaningful insights, ensuring that recommendations align with personal choices while benefiting from AI-driven product recommendations.

3.3.2 Skin Undertone Classification with Deep Learning

3.3.2.1 Convolutional Neural Network (CNN)

Once data is collected, wrist vein image processing is performed to classify skin undertones using deep learning techniques: CNN, Mobile Net V2 and DenseNet121. The first model, CNN model, for skin undertone classification is shown in Figure 3.5. The CNN structure employed for this experiment comprised five layers, including input, convolutional, max-pooling, dense, and output layers. CNN, a feedforward network, receives images as input at the input layer. Therefore, the wrist vein images were accepted at the input layer. The input layer was created to handle 128 x 128 x 3 images. And then a convolutional layer was implemented to select hierarchical features from the images. Following the convolutional layer, the feature maps are flattened and delivered through a fully connected layer. The output layer will be probabilities corresponding to each skin undertone category, such as warm, cool, and neutral.

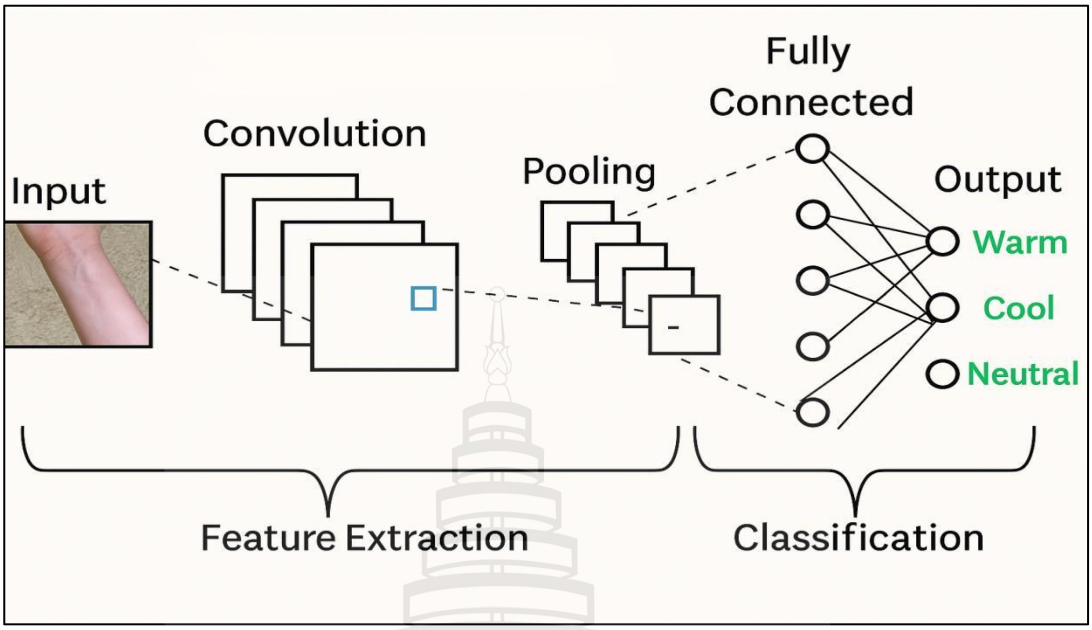


Figure 3.5 Skin Undertone Classification with CNN

3.3.2.2 Mobile Net V2

The second model, the Mobile Net model, is illustrated in Figure 3.6. An "Input Image" (depicting skin) first undergoes "Preprocessing" to standardize it to a 128x128 pixel resolution. This preprocessed image then enters the "MobileNetV2" backbone, which progressively extracts features through a sequence of convolutional layers with ReLU activation (shown by green arrows) and max pooling operations (shown by red arrows) to minimize spatial dimensions, shown as feature maps decreasing from 128x128 to 64x64, then 32x32, and finally to 4x4, while increasing the number of channels (n=32, n=96, n=1280). After that, the final 4x4 feature map is "Flattened" and then fed into a "fully connected" "classifier" layer, which has multiple neurons leading to a SoftMax activation function. This SoftMax layer outputs probabilities for the three skin undertone categories: cool, warm, and neutral.

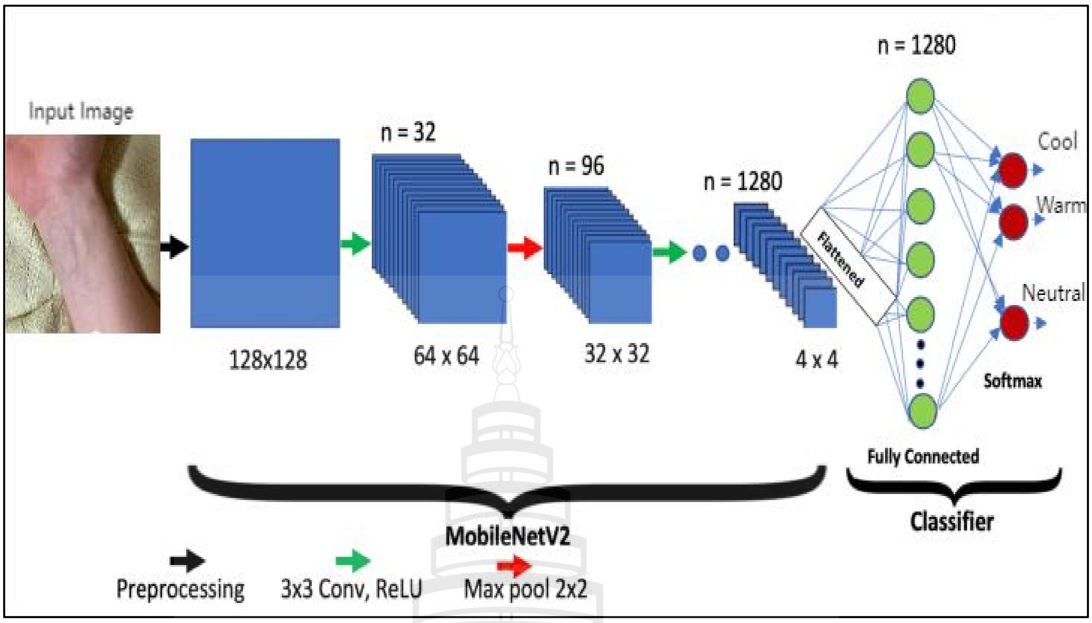


Figure 3.6 Skin Undertone Classification with Mobile Net V2

3.3.3 Dense Net121

The third model, DenseNet121 model, for skin undertone classification is designed to feed into a lipstick shade recommendation system. The model begins by processing an "Input Image" through initial "Convolution" and "Pooling" layers for feature extraction and dimension reduction. Its core consists of a series of four "Dense Blocks" and three "Transition Layers," a DenseNet121 architecture that promotes feature reuse and efficient information flow through concatenation. The extracted high-level features are then passed to a "Classification layer," which categorizes the skin undertone as cool, warm, and neutral. DenseNet121 model architecture is demonstrated in Figure 3.7.

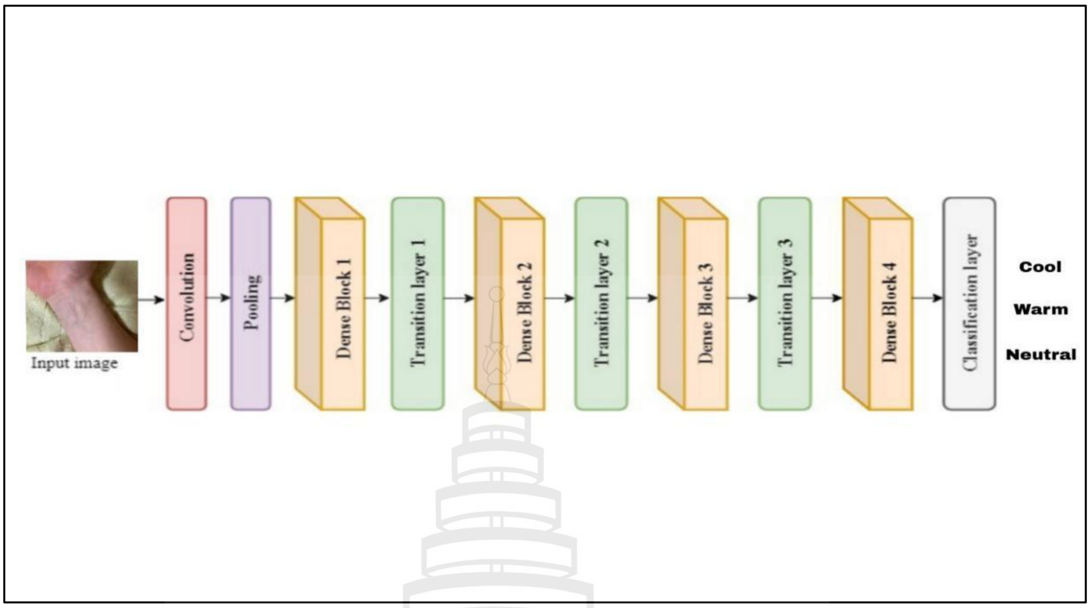


Figure 3.7 Skin Undertone Classification with DenseNet121

3.3.4 Lipstick Recommendation with Clustering Methods

In parallel, lipstick product information processing is executed using clustering methods to segment and recommend lipsticks products based on products' characteristics and user preferences. The initial phase of this clustering method begins by loading the lipstick product dataset and isolating key categorical features such as 'Undertone', 'Finish', 'Benefit', and 'Ethical Preferences'. These chosen attributes then are transformed into a numerical format using OneHotEncoder via a ColumnTransformer.

In this phase, K-Means, DBSCAN, and WOA K-Medoids methods are utilized to determine the best method for lipstick products recommendation. For K-Means and WOA K-Medoids, a range of K values is systematically tested by iterating from 2 up to a maximum of either 10 or the number of unique samples minus one to ensure the formation of meaningful clusters, with each configuration evaluated using the Silhouette Score, a metric that assesses clustering quality and cohesion. For DBSCAN, different combinations of eps (neighborhood radius) and min_samples (minimum points to form a dense cluster) are explored. Silhouette Score is used to find the best method of these three approaches. Ultimately, the clustering method with the highest Silhouette

Score—whether K-Means, K-Medoids, or DBSCAN—is selected as the most suitable method for the recommendation.

After identification of the optimal clustering method, user data is added to find the closest point with the cluster centroid of the original data clusters and the new user data to offer a recommendation. The new user data is calculated by taking the Euclidean distance from the centroid (mean) for each established cluster. The cluster whose centroid is closest to the new user is then chosen as the most relevant group. The top four recommended products are manufactured from the closest cluster. Finally, the top two products are filtered by price.

3.3.5 Lipstick Recommendation with Content-Based Filtering

The process of content-based lipstick product recommendation is started with the loading of the lipstick product dataset and the extraction of data features, including benefits, finishes, and ethical preferences. And then, features are transformed into numerical vectors by utilizing the TfidfVectorizer function, allowing a machine to understand the relative significance of words across the dataset. The shape of the resulting TF-IDF matrix is printed in Table 3.6, where each row represents a lipstick product and columns represent words extracted from the features.

Table 3.6 TF-IDF Product-Features Matrix

Product Name	Creamy	Features Hydrating	Vegan
B.O.M. Cloud Blur Lipstick	0.5010	0.2298	0.2544
Dior Addict Lip Maximizer	0.0000	0.2298	0.2544

Then, similarity between products is calculated with cosine similarity, which measures how close two items are based on their features. The result is stored in a similarity matrix and shown in Table 3.7. Overall processes are employed with the Python programming language along with necessary libraries like NumPy, Pandas, and sklearn.

Table 3.7 Consine Similarity Product Matrix for Lipstick Products

Product Name	Sheer Moisture Lip Tint	Long-Wearing Matte & Satin Velvet Lipstick	Speak for Yourself Hydrating Lipstick
B.O.M. Cloud Blur Lipstick	0.3799	0.3902	0.0558
Dior Addict Lip Maximizer	0.2212	0.3072	0.1222

CHAPTER 4

RESULT AND EVALUATION

4.1 Performance Evaluation for Content-Based Skincare Recommendations

To assess the performance of the content-based skincare recommendation, specialists' selection was considered. An online survey with an Excel file listing product names and ingredients from the dataset was conducted with five skincare specialists who are familiar with various skincare products. If the specialists select a product, it indicating that the specialists consider the products to be similar based on their ingredients. Three of the five skincare specialists must choose the recommended products from the content-based skincare recommendation. Accuracy is determined by dividing the entire number of matched cells, that is, cells that are both suggested and true in the ground truth, by the total number of true cells in the ground truth, considering these recommendations. The formula for accuracy evaluation is shown in equation 1, and its outcomes are provided in Table 4.1.

$$Accuracy = \frac{Total\ Matched\ Cells}{Total\ TRUE\ Cells} \tag{1}$$

Table 4.1 Performance Metric of Two Similarity Metrics

Product Categories	Accuracy based on Cosine Similarity	Accuracy based on Jaccard Similarity
Moisturizer	0.6250	0.5714
Cleanser	0.4286	0.5714
Toner	0.6667	0.3333
Serum	0.8000	0.8000
Sunscreen	0.8000	0.8000
Average	0.6640	0.6152

From Table 4.1, the result shows that the proposed method performed with an average accuracy of 66.40% with cosine similarity as well as 61.52% with Jaccard similarity across all product categories: moisturizer, cleanser, toner, serum, and sunscreen. This indicates that 66.40% of the recommended products matched the skincare specialists' selections when using cosine similarity and an average accuracy of 61.52% using Jaccard similarity. Examining the product categories individually, the proposed method achieved notably high accuracy for serum and sunscreen, with both metrics reaching an impressive 80% for these categories. This suggests that for products like serums and sunscreens, where ingredient compositions are typically more focused and specific, both similarity metrics are equally capable of accurately identifying suitable recommendations. Success in these categories reflects the proposed method's ability to effectively capture ingredient patterns and match them with skincare specialist selections. The overall findings indicate that the method outperformed utilizing the cosine similarity metric compared to the Jaccard similarity metric. Higher accuracy with cosine similarity suggests that it is more effective at capturing the nuances in ingredient similarity, leading to more accurate recommendations.

4.2 Performance Evaluation for Lipstick Recommendations

4.2.1 Evaluation of Deep Learning-based Skin Undertone Classification

The models were trained throughout a variety of epochs, specifically between 6 to 20 epochs, with a batch size of 32. After twenty epochs, the experiments showed no significant progress. After 20 epochs, the models' training and validation accuracy balance was optimum, indicating an excellent fit. As a result, 20 training epochs were chosen as the ending number. Figure 3.6, 3.7, and 3.8 shows the performance of the proposed structures for 20 epochs during the training and validation phases respectively.

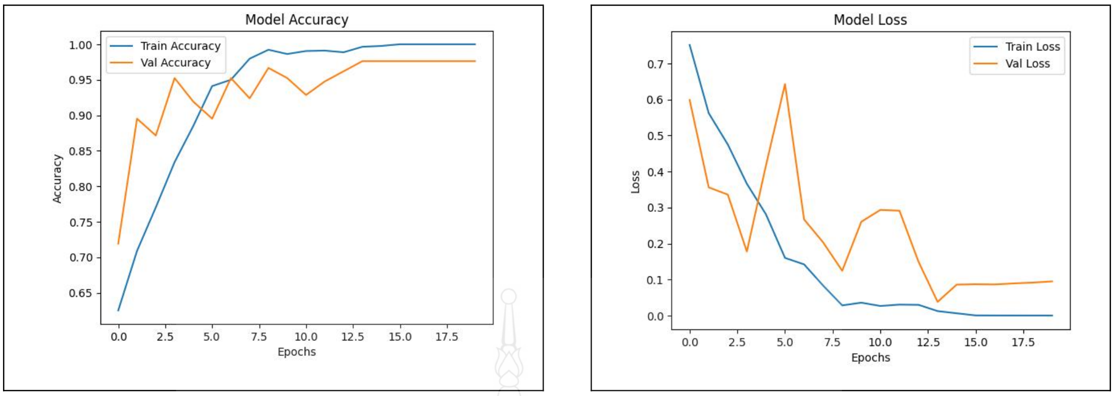


Figure 4.1 Accuracy and Loss for CNN (epochs = 20)

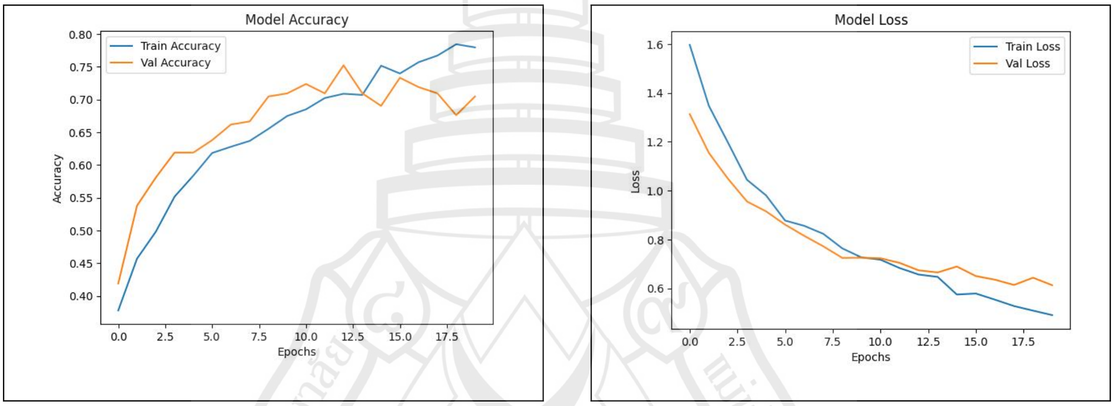


Figure 4.2 Accuracy and Loss for Mobile Net V2 (epochs = 20)

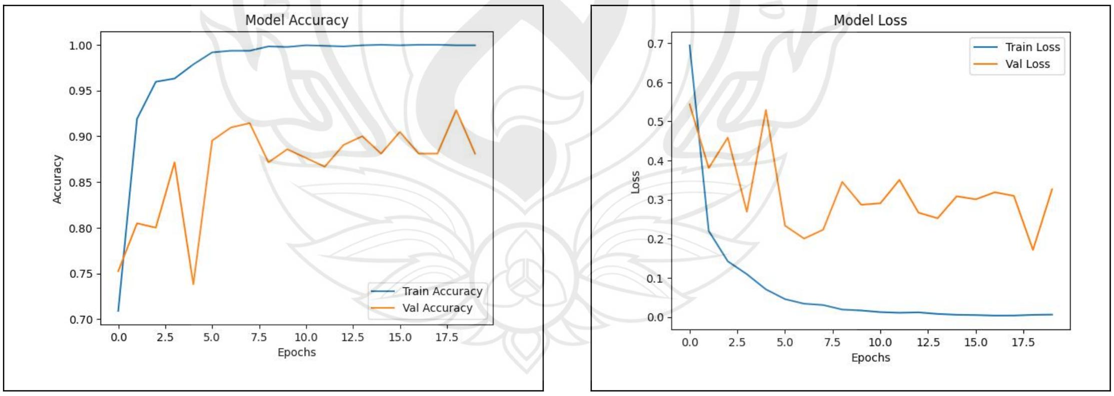


Figure 4.3 Accuracy and Loss for DenseNet121 (epochs = 20)

To measure the performance of the suggested models, accuracy, precision, recall, and F1-score were conducted. In the three models, the best accuracy rates for the entire dataset are as described in Table 4.2.

Table 4.2 Performance Metrics of Three Deep Learning Models

Model	Accuracy
CNN	0.84
Mobile Net V2	0.69
DenseNet121	0.77

Among the classification algorithms, CNN was the best performing classification algorithm, with an accurate score of 0.84 for the skin undertone dataset. Precise performance for each classification model is provided in Table 4.3, Table 4.4, and Table 4.5 respectively.

Table 4.3 Performance Metrics for CNN Model

Skin Undertone	Precision	Recall	F1-score
Cool	0.82	1.00	0.90
Warm	0.78	0.71	0.75
Neutral	0.92	0.80	0.85

Table 4.4 Performance Metrics for Mobile Net V2 Model

Skin Undertone	Precision	Recall	F1-score
Cool	0.71	0.86	0.78
Warm	0.57	0.71	0.64
Neutral	0.90	0.50	0.64

Table 4.5 Performance Metrics for DenseNet121 Model

Skin Undertone	Precision	Recall	F1-score
Cool	0.93	0.96	0.94
Warm	0.63	0.91	0.58
Neutral	0.86	0.44	0.74

The precision metric quantifies the percentage of affirmative identifications that were indeed accurate. A low false positive rate is indicated by a high precision, which means that the model is highly likely to be right when it predicts a particular undertone (such “Cool”). For example, in the third model, the “Cool” class has a precision of 0.93, indicating that 93% of the instances predicted as “Cool” were indeed “Cool.” Conversely, a low precision, like the 0.57 for “Warm” in the second model, suggests a greater rate for false positives, meaning many instances predicted as “Warm” were actual other undertones.

The percentage of true positives which were accurately observed is called recall, sometimes referred to as true positive rate or sensitivity. When a model owns a high recall and a low false negative rate, it is effective in identifying every instance of a particular class. For instance, the first model’s “Cool” recall of 1.00 means it correctly identified all actual “Cool” undertones. In contrast, the “Neutral” class in the third model gets a low recall of 0.44, showing that it missed a significant portion of the actual "Neutral" undertones.

The F1-score provides a single statistic that balances precision and recall by taking the harmonic means of both. A high F1-score suggests that the model is good at correctly recognizing positive cases and avoiding false positives and negatives, indicating that it performs well in both precision and recall. For example, the “Cool” class in the third model has a high F1-score 0.94, demonstrating a strong balance between precision and recall. Conversely, a low F1-score, such as the 0.58 for “Warm” in the third model, indicates an imbalance or poor performance in either precision or recall, or both, highlighting a trade-off where one metric might be good while the other is poor, or both are moderate.

In this study, the models generally perform well on the “Cool” undertone, consistently achieving high precision, recall, and F1-scores. Achievement for “Neutral” and especially “Warm” undertones is more varied, with some models showing strong precision but lower recall (e.g., “Neutral” in the second model) or vice-versa (e.g., “Warm” in the third model), leading to more moderate F1-scores. This suggests that while the models are adept at recognizing “Cool” undertones, there’s more variability and room for improvement in distinguishing “Neutral” and “Warm” undertones, particularly in avoiding misclassifications that impact recall.

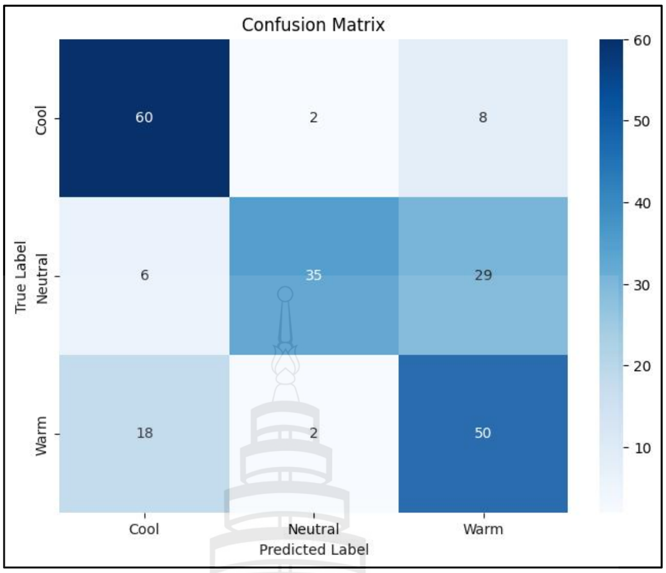


Figure 4.4 Confusion Matrix for Skin Undertone Classification with CNN

The confusion matrixes were visibly portrayed by the off-diagonal elements in Figures 4.4, 4.5, and 4.6, providing an in-depth overview of the performance of CNN, Mobile Net V2, and DenseNet121 for skin tone classification model across "Cool," "Neutral," and "Warm" classes. The CNN model shows high accuracy for "Cool" (70 correct, 0 misclassified) and "Neutral" (56 correct, 0 misclassified as Cool), but a notable misclassification of 15 "Warm" instances as "Cool" and 5 "Neutral" instances as "Warm".

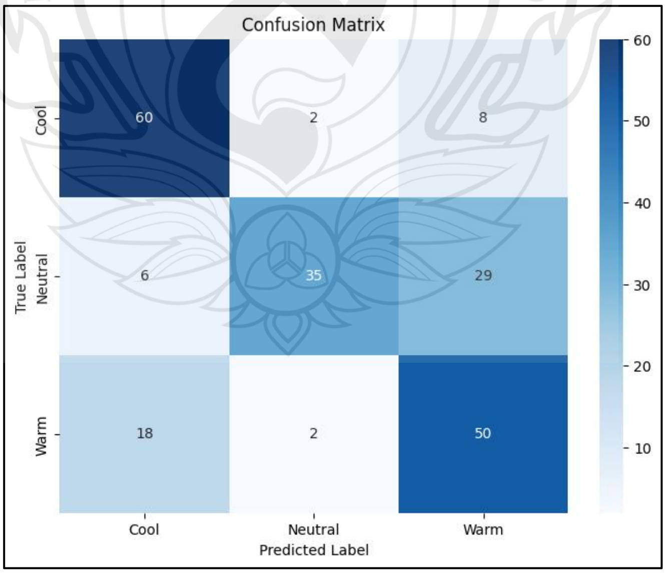


Figure 4.5 Confusion Matrix for Skin Undertone Classification with Mobile Net V2

Mobile Net V2 model indicates 60 correct “Cool” classifications with some misclassifications to “Neutral” (2) and “Warm” (8), while “Neutral” had 35 correct but 6 misclassified as “Cool” and 29 as “Warm”. “Warm” had 50 correct classifications, with 18 misclassified as “Cool” and 2 as “Neutral”.

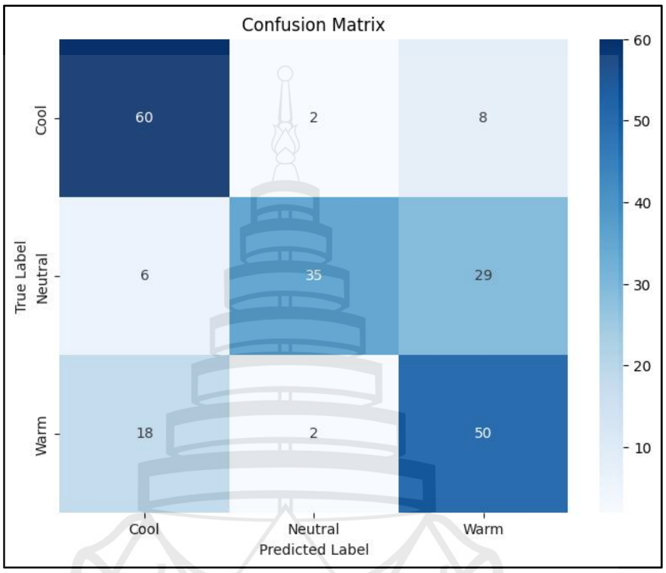


Figure 4.6 Confusion Matrix for Skin Undertone Classification with DenseNet121

DenseNet121 demonstrates 67 correct “Cool” predictions with minimal misclassifications (3 to “Warm”), and 31 correct “Neutral” predictions with significant misclassifications to “Warm” (35) and fewer to “Cool” (4). “Warm” instances were largely correctly classified (64), with minor misclassifications to “Cool” (1) and “Neutral” (5).

The explored findings for skin undertone classification using three deep learning models: CNN, MobileNetV2, and DenseNet121 showed that 20 epochs provided the best training and validation accuracy. The CNN model outperformed MobileNetV2 (0.69) and DenseNet121 (0.77), scoring an overall accuracy of 0.84. While all models performed well for “cool” undertones in terms of precision, recall, and F1-scores, their ability to categorize “neutral” and “warm” undertones differed, with confusion matrices indicating instances of misclassification, notably for “warm” undertones. Despite these limitations, the CNN model continually performed better. The best-performing model's accurate skin undertone classifications were then used as input for subsequent recommendation algorithms, which recommended appropriate lipstick shades based on product attributes and use preferences.

4.2.2 Performance Evaluation for Lipstick Recommendations

Performance for lipstick recommendations approaches are evaluated as follows:

4.2.2.1 Clustering-Based Lipstick Recommendations

This section provides a comprehensive evaluation of the clustering algorithms employed to categorize lipstick shades according to product characteristics and user preferences. The performance of three distinct clustering algorithms- K-Means, DBSCAN, and WOA K-Medoids was measured with the Silhouette Index Score. The Silhouette Score, which compares an object's similarity to its own cluster to other clusters, is a commonly used statistic for assessing the quality of clusters. Clusters that are dense and well-separated are indicated by high scores (closer to +1), which show that objects are well-matched to their own cluster and poorly matched to surrounding clusters. On the other hand, negative values imply that data points may have been allocated to the incorrect clusters, whereas values close to 0 suggest overlapping clusters.

To determine the best clustering algorithm, the dataset was divided into three skin undertone categories: “Cool,” “Warm,” and “Neutral.” The performance was evaluated using Silhouette Scores, as shown in Table 4.6. DBSCAN consistently outperformed K-Means and WOA K-Medoids, achieving the highest scores for both the “Cool” (0.96) and “Warm” (0.91) undertones. In the “Neutral” group, although K-Means scored slightly higher at 0.86, DBSCAN still showed competitive performance with a score of 0.74, significantly surpassing WOA K-Medoids (0.44).

Table 4.6 Silhouette Score Comparison of Three Clustering Methods

Lipstick Dataset with Skin Undertone	K-Means	DBSCAN	WOA K-Medoids
Cool	0.93	0.96	0.94
Warm	0.63	0.91	0.58
Neutral	0.86	0.74	0.44

According to these results, DBSCAN consistently emerged as the superior clustering algorithm across all skin undertone categories demonstrate its superiority in effectively grouping lipstick shades. This indicates that DBSCAN produced reasonably well-separated and cohesive clusters. The consistently lower scores for K-Means and WOA K-Medoids suggest their struggle in effectively categorizing lipstick shades according to undertone. Specifically, DBSCAN, with optimized parameters of epsilon (eps) = 4 and minimum samples (min-sample) = 2, proved to be the optimal choice for grouping the diverse lipstick product data, as evidenced by its superior Silhouette Scores for “Cool” (0.96), “Warm” (0.91), and “Neutral” (0.74) lipstick datasets.

4.2.2.2 Nearest Neighbor Recommendation with New User Data

Following the identification of the optimal DBSCAN clustering model, the recommendation process integrates new user data. The clustering results using DBSCAN with new user data integration are demonstrated as Figure 4.7. Different colors denote clusters, numbered from 0 to 7, while black points signify noise (-1), which the model couldn’t assign to a defined cluster. A new user is marked by a red ‘X,’ indicating their preferences were mapped onto the existing clusters. Based on the DBSCAN parameters (epsilon = 4.0, min_samples = 2), this user has been placed in Cluster 7, suggesting that their attributes align closely with the characteristics of products within this cluster. The dashed connection between the new user and Cluster 7 visually illustrates their association, reinforcing the personalized recommendation process. This clustering approach enables more accurate user segmentation, improving lipstick recommendations tailored to individual preferences.

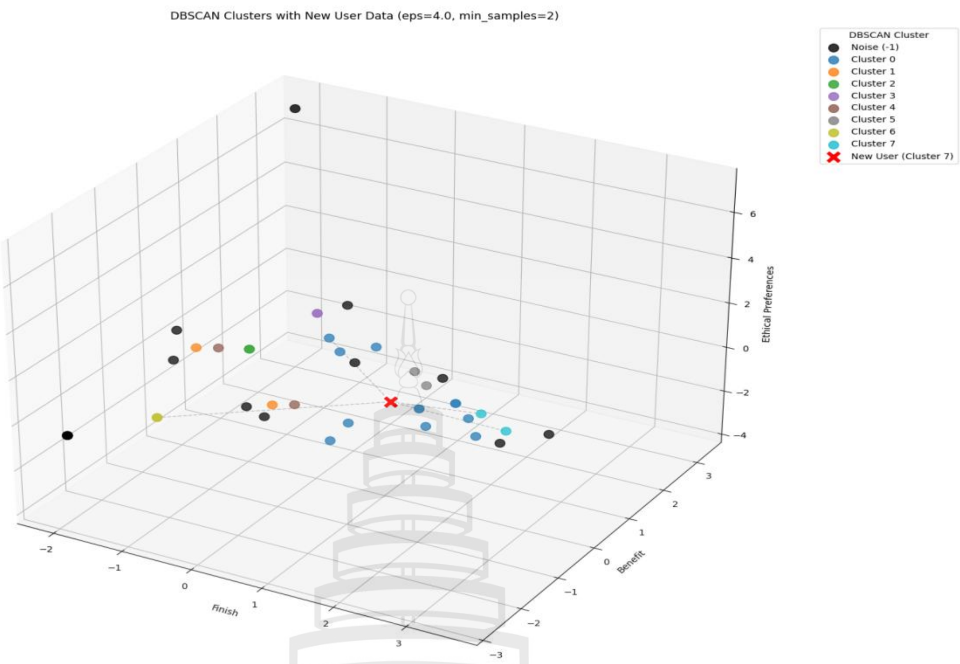


Figure 4.7 Nearest Neighbor Recommendation with New User Data

4.2.2.3 Clustering with Nearest Centroid Recommendation

Afterwards, the Euclidean distance between the new user's preferences and the centroid (mean) of every DBSCAN cluster was calculated. The cluster whose centroid is nearest to the new user's data point is then identified as the most relevant group for recommendation. This process is illustrated in the provided 3D visualization (Figure 4.8), where a “New User” (marked with a red ‘X’) is added to the existing DBSCAN clusters. The dotted magenta line visually represents the Euclidean distance calculation, clearly showing the new user's proximity to the centroid of ‘Cluster 7’ (marked with a light green ‘X’).

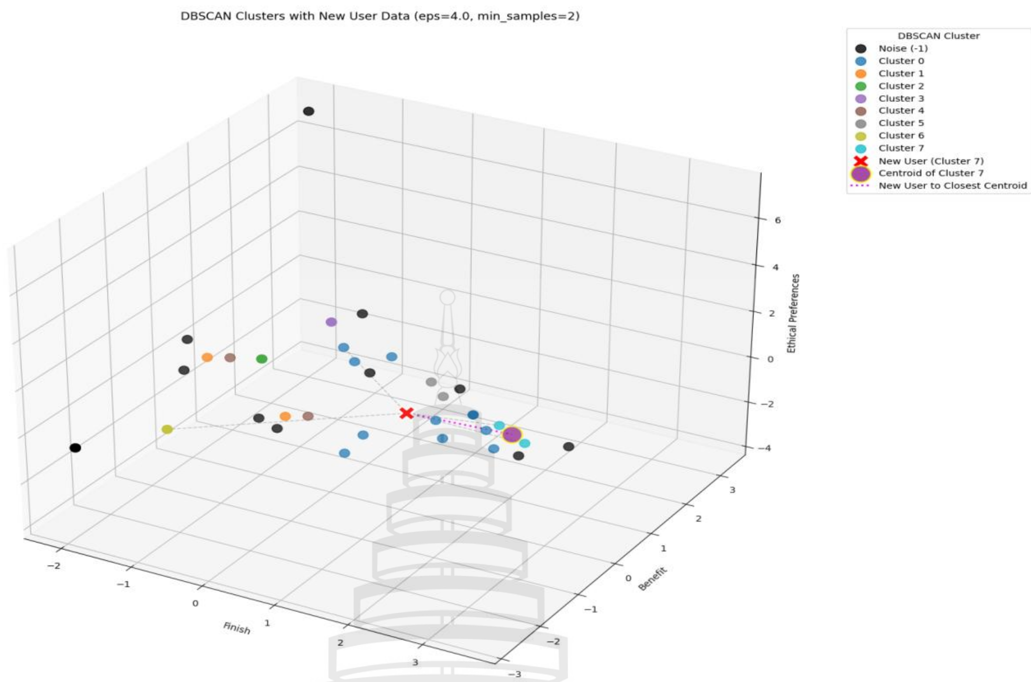


Figure 4.8 Clustering with Nearest Centroid Recommendation with New User Data

Once the closest cluster is identified, the system recommends the top four products from within this cluster. These four products are chosen based on their proximity to the cluster's centroid, ensuring that they are representative of the user's inferred preferences. To refine the recommendations further and enhance user satisfaction, a final filtering step is applied, where the top two products are selected based on their price. As demonstrated in Figure 4.9, after a new user's closest cluster is identified as ‘Cluster 7’, the top two recommended products within this cluster are “Top Product: Mineral Fusion (Rich Plum)” (green circle) and “Top Product: Almay (Berry Light)” (light green circle)—were highlighted.

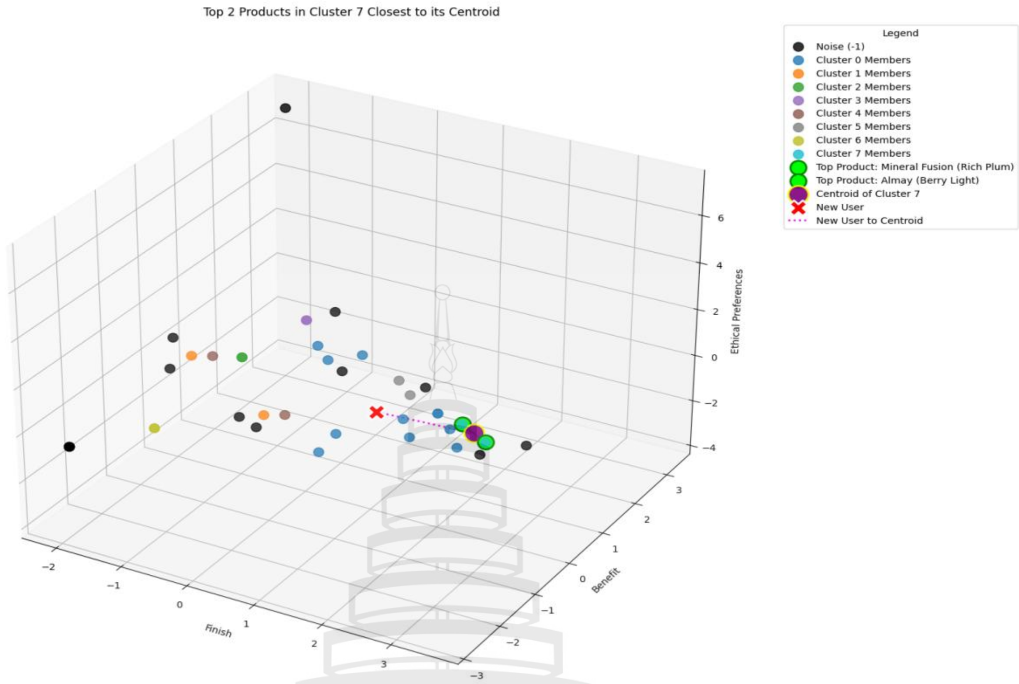


Figure 4.9 DBSCAN Clustering for Lipstick Recommendation

4.2.3 Content-Based Lipstick Recommendations

Based on the content-based approach, an example lipstick ("B.O.M Cloud Blur Lipstick", Cool, Creamy Finish, Hydrating, Vegan) and similar recommended products are described in Table 4.7.

Table 4.7 Sample of Content-Based Recommendations

Product Name	Undertone	Finish	Benefit	Ethical Preferences
Sheer Moisture Lip Tint	Cool	Creamy	Hydrating	Vegan, Cruelty-free
Long-Wearing Matte & Satin Velvet Lipstick	Cool	Matte	Long-wearing	Cruelty-free, Alcohol-free, Oil-free
Dramatically Different™ Lipstick Shaping Lip Color	Cool	Creamy	Hydrating	Vegan, Parabens free, Phthalates free
Weightless Lip Color Nourishing Satin Lipstick	Cool	Satin	Hydrating, Long-wearing	Cruelty-free

4.2.4 Performance Evaluation Comparisons

For the evaluation of lipstick recommendations, a rigorous comparative analysis was conducted across 30 users, directly comparing the performance of clustering-based (specifically using DBSCAN) and content-based recommendation approaches. An expert generated two recommended lipstick products for each user from each approach, resulting in a total of six candidate lipsticks per user. The expert then meticulously evaluated these six options, selecting the top two choices in order of preference, considering the lipstick's properties, price, and congruence with the participant's individual skin undertone.

A quantitative scoring system was implemented to assess the accuracy of each approach. A score of 1 was awarded if a recommended lipstick matched both the expert's chosen name and its assigned order (first or second preference), indicating a perfect match in relevance and ranking. A partial score of 0.5 was given if only the name matched but the order was incorrect, acknowledging relevance but penalizing for imperfect ranking. Finally, a score of 0 was assigned if there was no match. The detailed example of this evaluation process, exemplified with B.O.M Cloud Blur Lipstick, is described in Table 4.8.

Table 4.8 Evaluation Measures for Lipstick Recommendations

Recommended Products	Expert's Evaluation		
	Name	**Order**	**Score**
Sheer Moisture Lip Tint (Clustering-Based Recommendation)	Yes	Yes	1
Almay Smart Shade Butter Kiss Lipstick (Clustering-Based Recommendation)	No	No	0
Almay Smart Shade Butter Kiss Lipstick (Clustering-Based Recommendation with Centroid Analysis)	No	No	0
Sheer Moisture Lip Tint (Clustering-Based Recommendation with Centroid Analysis)	Yes	No	0.5
Sheer Moisture Lip Tint (Content-Based Recommendation)	Yes	Yes	1
Long-Wearing Matte & Satin Velvet Lipstick (Content-Based Recommendation)	Yes	Yes	1

For this specific example, the content-based recommendation successfully achieved a perfect score of 1 for “Sheer Moisture Lip Tint” (matching both name and first preference) and another score of 1 for “Long-Wearing Matte & Satin Velvet Lipstick” (matching name and second preference), demonstrating its strong performance in both relevance and order. In contrast, the clustering-based approach yielded a score of 1 for “Sheer Moisture Lip Tint” but 0 for “Almay Smart Shade Butter Kiss Lipstick,” while the clustering-based with centroid analysis received 0.5 for “Sheer Moisture Lip Tint” (due to a name match but incorrect order) and 0 for “Almay Smart Shade Butter Kiss Lipstick.”

The aggregated accuracy results across all 30 users are detailed in Table 4.9, differentiating between “with order” and “without order” accuracy. “With order” accuracy, a more stringent metric, reflects the system’s ability to provide both the correct product and its exact preferred ranking. “Without order” accuracy, conversely, focuses solely on whether the recommended products were relevant, regardless of their position within the top two.

Table 4.9 Evaluation Comparisons for Lipstick Recommendations

Proposed Method	**Accuracy**	
	With Order	**Without Order**
Content-Based Filtering	76.6%	83%
Nearest Neighbor Recommendation	66.7%	73%
Clustering with Nearest Centroid Recommendation	40%	48%

The results unequivocally show that the content-based filtering approach emerged with the highest overall accuracy. It achieved an impressive 76.6% accuracy when considering the order of preference (“with order”) and an even higher 83% accuracy when only considering the relevance of the recommended products (“without order”). This indicates its superior ability to generate recommendations that align more closely with expert judgment for personalized lipstick selection, both in terms of identifying the right products and ranking them correctly.

The Clustering-Based Recommendation approach followed, demonstrating a respectable 66.7% accuracy “with order” and 73% “without order.” While performing adequately, it lagged the content-based method, particularly in its ability to predict the precise preference order.

Conversely, the Clustering-Based Recommendation with Centroid Analysis exhibited the lowest performance among the three methods, achieving only 40% accuracy “with order” and 48% “without order.” This suggests that while incorporating centroid analysis into the clustering approach might aim for improved recommendations, in this specific context, it did not enhance the accuracy of personalized lipstick suggestions to the same extent as the other methods.

In conclusion, the content-based filtering approach proved to be the most effective method for personalized lipstick recommendations in this study, consistently outperforming both clustering-based methods in accurately identifying and ranking preferred products according to expert judgment.

4.2.5 Computing Complexity

The computational performance of recommendation systems is a critical factor, with significant differences observed between content-based filtering and clustering-based approaches. Content-based filtering, which primarily relies on matching item features with a user's historical preferences, often involves calculating similarity scores (e.g., cosine similarity, Jaccard similarity) between a target item and all items in a user's profile, or between a new item and all existing items. For large datasets, this can lead to substantial computational costs, as the process typically involves iterating through a considerable number of items and their associated features. The complexity of content-based filtering can scale with the number of items and the dimensionality of their feature vectors, potentially becoming a bottleneck in systems with a vast product catalog or highly descriptive attributes, particularly during the initial processing of item descriptions using methods like TF-IDF.

Conversely, clustering-based recommendation systems aim to improve computational efficiency by pre-processing user or item data into groups. By segmenting users with similar preferences or items with similar attributes into clusters, the recommendation process can be significantly streamlined. Instead of comparing a user or item against the entire dataset, computations are confined to a much smaller subset of relevant clusters. For instance, in user-based clustering, once a user's cluster is identified, recommendations are generated only from items popular within that specific cluster. While the initial clustering phase can be computationally intensive, especially for algorithms that scale quadratically with the number of data points (e.g., some hierarchical clustering methods), the overhead is typically amortized over numerous recommendation requests. Thus, for very large and dynamic datasets, clustering-based approaches can offer superior scalability and faster real-time recommendations once the clusters are established and maintained.

CHAPTER 5

CONCLUSION AND DISCUSSION

5.1 Research Conclusion

This research presented a comprehensive approach to cosmetic product recommendation by integrating multiple methodologies tailored for skincare and lipstick selection. The study collected product data from e-commerce platforms and wrist images data from Roboflow website. By employing content-based filtering for skincare recommendations based on ingredient profiles, deep learning in conjunction with clustering, and content-based approaches for lipstick recommendations, product appropriateness is enhanced based on individual skin undertones. The study evaluated various similarity measures (cosine similarity and Jaccard similarity) for content-based skincare recommendations to guarantee appropriate ingredient matching. Deep learning models: CNN, MobileNetV2, and DenseNet121 are applied to classify skin undertones for lipstick recommendation, while clustering approaches (K-Means, DBSCAN, and WOA-K Medoids) refine the recommendation process. Model performance is evaluated by employing accuracy, precision, recall, and F1-score, illustrating the efficacy of various approaches in improving suggestion accuracy. Furthermore, experts review is provided to provide qualitative insights into the recommendations, proving their suitability in professional settings.

5.2 Research Discussion

This research successfully demonstrated the efficacy of a tailored, multi-faceted approach to cosmetic product recommendation, addressing the distinct challenges posed by both skincare and lipstick selection. The skincare recommendation utilized a dataset of 45 products (5 features) from e-commerce sources. For skincare recommendations, the study would suggest implementing content-based filtering, leveraging cosine similarity along with expert evaluation to effectively match products to user needs based on detailed ingredient profiles. Simultaneously, for lipstick recommendations, the study would recommend a two-stage process: initially, using the CNN deep learning model for robust skin undertone classification due to its superior accuracy (84%). The categorized skin undertone output would then be combined with content-based filtering for a lipstick product dataset consisting of 117 lipstick products (9 characteristics) to align suggestions with user preferences, which would be enhanced further using expert feedback. The combined techniques highlight a huge step forward in personalized beauty technology, providing practical solutions that can reduce customer aggravation, enhance purchasing decisions, and build a more educated interaction between consumers and cosmetic items.

5.3 Research Limitations and Future Works

The current approaches primarily focus on static ingredient profiles for skincare and skin undertone classification for lipstick, which may not fully capture the dynamic beauty trends. The scope of the recommendation is also confined to skincare and lipstick, leaving other cosmetic categories unexplored. Furthermore, the quantity and accessibility of datasets allows it challenging to classify skin undertones to provide product recommendations.

Future studies will be explored the dynamic adaptation of recommendations based on real-time user feedback and evolving beauty trends. Futures studies will also incorporate advanced deep learning models for feature extraction from product images for both skincare and lipstick can further refine content-based filtering. Further experiments will be extending these frameworks to other cosmetic categories, such as haircare or makeup, and developing hybrid models that blend collaborative filtering with the content-based and deep learning approaches, presents exciting avenues for continued innovation.

REFERENCES

Abdurrafi, M. F., & Ningsih, D. H. U. (2023). Content-based filtering using cosine similarity algorithm for alternative selection on training programs. *Journal of Soft Computing Exploration, 4*(4), 4. https://doi.org/10.52465/joscex.v4i4.232

Albelwi, S. A. (2022). Deep architecture based on DenseNet-121 model for weather image recognition. *International Journal of Advanced Computer Science and Applications*, *13*(10). https://doi.org/10.14569/IJACSA.2022.0131065

Amin, M. F. (2023). Confusion matrix in three-class classification problems: A step-by-step tutorial. *Journal of Engineering Research, 7*(1), 26.

Ariawan, K. R., Ekayana, A. A. G., Indrawan, I. P. Y., Winatha, K. R., & Setiawan, I. N. A. F. (2025). Performance comparasion of DenseNet121 and MobileNetV2 for cacao fruit disease image classification. *Indonesian Journal of Data and Science*, *6*(1), 1. https://doi.org/10.56705/ijodas.v6i1.233

Berg, A., Hudson, S., Weaver, K. K., Pacchia, M. L., & Amed, I. (2023). *The beauty market in 2023: A special State of Fashion report*. https://www.mckinsey.com/industries/retail/our-insights/the-beauty-market-in-2023-a-special-state-of-fashion-report

Britannica. (n.d.). *Cosmetic*. Retrieved November 24, 2024, from https://www.britannica.com/art/cosmetic

Cahyono, C., Jonatan, S., Tan, W. S., Pratama, G. D., & Hasana, S. (2023). Using convolutional neural network to improve decision making in cosmetic purchases. In *2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*. IEEE. https://doi.org/10.1109/AiDAS60501.2023.10284688

Cartwright, M. (n.d.). *Cosmetics in the ancient world*. World History Encyclopedia. Retrieved November 24, 2024, from https://www.worldhistory.org/article/1441/cosmetics-in-the-ancient-world/

Chen, Y., Dai, X., Chen, D., Liu, M., Dong, X., Yuan, L., . . . Liu, Z. (2022). *Mobile-former: Bridging MobileNet and transformer* (arXiv:2108.05895). https://doi.org/10.48550/arXiv.2108.05895

Chenan, H., & Tsutsumida, N. (2025). *A scalable K-medoids clustering via whale optimization algorithm* (arXiv:2408.16993). https://doi.org/10.48550/arXiv.2408.16993

Cheng, D., Zhang, C., Li, Y., Xia, S., Wang, G., Huang, J., . . . Xie, J. (2024). GB-DBSCAN: A fast granular-ball based DBSCAN clustering algorithm. *Information Sciences*, *674*, 120731. https://doi.org/10.1016/j.ins.2024.120731

Consumer Insights. (2017). Retrieved November 23, 2024, from https://www.cosmeticseurope.eu/files/6114/9738/2777/CE_Consumer_Insights_2017.pdf

Fahad Iqbal, T., & Gnanajeyaraman, R. (2023). Hybrid content-collaborative filtering: Improving collaborative filtering and neural network-based recommendation. In *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE. https://doi.org/10.1109/ICSES60034.2023.10465386

Fayyadhila, R., Junaidi, A., & Prasetyo, N. A. (2021). Implementasi deep learning untuk klasifikasi citra undertone menggunakan algoritma convolutional neural network. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, *1*(2), 2. https://doi.org/10.20895/dinda.v1i2.366

Fortuna, D., Tahyudin, I., & Utomo, F. S. (2024). Optimization in makeup recommendations using RGB and YCbCr analysis with CNN algorithm for skin undertone. In *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. IEEE. https://doi.org/10.1109/ICITISEE63424.2024.10730153

Grand View Research. (2024). *Cosmetics market size, share and growth report, 2030*. https://www.grandviewresearch.com/industry-analysis/cosmetics-market

Herdiana, I., Alfin Kamal, M., Triyani, Nur Estri, M., & Renny. (2025). *A more precise elbow method for optimum K-means clustering* (arXiv:2502.00851). https://doi.org/10.48550/arXiv.2502.00851

Hu, L., Jiang, M., Dong, J., Liu, X., & He, Z. (2024). *Interpretable clustering: A survey* (arXiv:2409.00743; Version 1). https://doi.org/10.48550/arXiv.2409.00743

Itten, J., Haagen, E. van, & Itten, J. (1996). *The art of color: The subjective experience and objective rationale of color* (21. Aufl). Van Nostrand Reinhold.

Iwabuchi, R., Nakajima, Y., Honma, H., Aoshima, H., Kobayashi, A., Akiba, T., . . . Masuyama, S. (2017). Proposal of recommender system based on user evaluation and cosmetic ingredients. In *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*. IEEE. https://doi.org/10.1109/ICAICTA.2017.8090967

Jain, A., Jain, A., Chauhan, N., Singh, V., & Thakur, N. (2017). Information retrieval using cosine and jaccard similarity measures in vector space model. *International Journal of Computer Applications, 164*(6), 28-30.

Khatreja, A., & Mulay, P. (2022). Lipstick color suggested using hybrid hidden Markov model. *Cardiometry*, *22*, 421-428. https://doi.org/10.18137/cardiometry.2022.22.421428

Korichi, R., Pelle-De-Queral, D., Gazano, G., & Aubert, A. (2009). Why women use makeup: Implication of psychological traits in makeup functions. *J. Cosmet. Sci., 59*(2), 127-137.

Kumar, N., Yadav, S. K., & Yadav, D. S. (2020). Similarity measure approaches applied in text document clustering for information retrieval. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. IEEE. https://doi.org/10.1109/PDGC50313.2020.9315851

Laamouri, A., & Nawal, S. (2024). Image classification for visual recommendation using deep learning. In *Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security* (pp. 1-6). https://doi.org/10.1145/3659677.3659693

Lee, G., Jiang, X., & Parde, N. (2023). A content-based skincare product recommendation system. In *2023 International Conference on Machine Learning and Applications (ICMLA)*. IEEE. https://doi.org/10.1109/ICMLA58977.2023.00308

Liu, Y., Zeng, Y., Li, R., Zhu, X., Zhang, Y., Li, W., . . . Hu, G. (2024). A random particle swarm optimization based on cosine similarity for global optimization and classification problems. *Biomimetics, 9*(4), 204. https://doi.org/10.3390/biomimetics9040204

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

Lumintu, I. (2023). Content-based recommendation engine using term frequency-inverse document frequency vectorization and cosine similarity: A case study. In *2023 IEEE 9th Information Technology International Seminar (ITIS)*. IEEE. https://doi.org/10.1109/ITIS59651.2023.10420137

Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies, 12*(2), 2. https://doi.org/10.3390/technologies12020015

Mouhiha, M., Oualhaj, O. A., & Mabrouk, A. (2024). Combining collaborative filtering and content based filtering for recommendation systems. In *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*. IEEE. https://doi.org/10.1109/WINCOM62286.2024.10658169

Patty, J. C., Kirana, E. T., & Giri, M. S. D. K. (2018). Recommendations system for purchase of cosmetics using content- based filtering. *International Journal of Computer Engineering and Information Technology, 10*(1), 1–5.

Permana, A. H. J. P. J., & Wibowo, A. T. (2023). Movie recommendation system based on synopsis using content-based filtering with TF-IDF and cosine similarity. *International Journal on Information and Communication Technology (IJoICT), 9*(2), 1-14. https://doi.org/10.21108/ijoict.v9i2.747

Raihan, A., Ahmad Ibrahim, A. M., & Gozali, A. A. (2024). Hangout places recommendation system using content-based filtering and cosine similarity methods. *Journal of Dinda: Data Science, Information Technology, and Data Analytics, 4*(2), 2. https://doi.org/10.20895/dinda.v4i2.1464

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2019). *MobileNetV2: Inverted Residuals and Linear Bottlenecks* (arXiv:1801.04381). https://doi.org/10.48550/arXiv.1801.04381

Scott, W. (2021, September 26). *TF-IDF for document ranking from scratch in python on real world dataset*. https://towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089

Sexton, M. R. (2022). *Perspectives on retail and consumer goods issue 8*. http://archive.org/details/perspectives-on-retail-and-consumer-goods_issue-8

Shah, K., Salunke, A., Dongare, S., & Antala, K. (2017). Recommender systems: An overview of different approaches to recommendations. In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. IEEE. https://doi.org/10.1109/ICIIECS.2017.8276172

Sirisayan, P. (2022). *Personal color analysis based on color harmony for skin tone*. Chulalongkorn University Theses and Dissertations (Chula ETD). https://doi.org/10.58837/CHULA.THE.2022.191

Sukestiyarno, Y., Sapolo, H., & Sofyan, H. (2023). *Application of recommendation system on e-learning platform using content-based filtering with Jaccard similarity and cosine similarity algorithms*. Preprints.org. https://doi.org/10.20944/preprints202306.1672.v1

Suvarna, B., & Balakrishna, S. (2022). An efficient fashion recommendation system using a deep CNN model. In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*. IEEE. https://doi.org/10.1109/ICACRS55517.2022.10029063

Tare, S. S., Bhute, M. M., & Arage, P. (2023). Recevent: NLP based event recommender system. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. IEEE. https://doi.org/10.1109/ICAAIC56838.2023.10140251

Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A., Chavez-Urbiola, E. A., & Romero-Gonzalez, J. A. (2025). Loss functions and metrics in deep learning. *Artificial Intelligence Review*, *58*(7), 195. https://doi.org/10.1007/s10462-025-11198-7

Yin, H., Aryani, A., Petrie, S., Nambissan, A., Astudillo, A., & Cao, S. (2024). *A rapid review of clustering algorithms* (arXiv:2401.07389). https://doi.org/10.48550/arXiv.2401.07389

Ziogas, I. P., Streviniotis, E., Papadakis, H., & Chalkiadakis, G. (2022). Content-based recommendations using similarity distance measures with application in the tourism domain. In *Proceedings of the 12th Hellenic Conference on Artificial Intelligence* (pp. 1–10). https://doi.org/10.1145/3549737.3549772

CURRICULUM VITAE

NAME Theint Zar Lwin Kyaw

EDUCATION BACKGROUND

2020 Bachelor of Science in Business Information System

University of Information Technology, Myanmar

WORK EXPERIENCE

2021-2025 Business Development Executive

Myanmar Technology Gateway Company Ltd.

PUBLICATION

Kyaw, T. Z. L., Uttama, S., & Panwong, P. (2024). Leveraging ingredient profiles in Content-based skincare product recommendation. *In* 2024 8th International Conference on Information Technology (InCIT) (pp. 319–324). IEEE. https://doi.org/10.1109/InCIT63192.2024.10810620