Chatbot – Beauty Skin Care Products Recommendations

By

Liew Yi Kei

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONOURS) BUSINESS INFORMATION SYSTEMS

Faculty of Information and Communication Technology

(Kampar Campus)

JAN 2021

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: **Chatbot – Beauty Skin Care Products Recommendations**

Academic Session: __Y3S2___________

I LIEW YI KEI

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature) (Supervisor's signature)

Address:
29, Lorong 5, Taman Lake View.
34000, Taiping, Perak.

Dr. Aun Yichiet

Date: 21/4/22 **Date**: 21/4/22

FYP THESIS SUBMISSION FORM

Universiti Tunku Abdul Rahman			
Form Title : **Sample of Submission Sheet for FYP/Dissertation/Thesis**			
Form Number: **FM-IAD-004**	Rev No.: **0**	Effective Date: **21 JUNE 2011**	Page No.: **1 of 1**

FACULTY OF INFORMATION AND COMMUNNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 21/4/22

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that Liew Yi Kei (ID No: 1 8 A C B 0 2 0 5) has completed this final year project entitled "Chatbot - Beauty Skin Care Products Recommendations" under the supervision of Dr. Aun YiChiet (Supervisor) from the Department of Computer and Communication Technology, Faculty of Information and Communication Technology.

I understand that University will upload softcopy of my final year project in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

Liew Yi Kei

DECLARATION OF ORIGINALITY

I declare that this report entitled **Chatbot – Beauty Skin Care Products Recommendations** is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature :

Name : Liew Yi Kei

Date : 1/8/2021

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisor, Aun Yichiet who has given me this bright opportunity to engage in this project. A million thanks to you. I am also grateful and appreciate the support from my friends.

Finally, I must say thanks to my parents and my family for their love, support, and continuous encouragement throughout the course.

ABSTRACT

Skincare products work differently depending on the type of skin. Therefore, the project proposes a context-aware chatbot for skincare product recommendations based on skin types. Firstly, we collect genuine product reviews dataset using a custom web crawler on cosmetic websites. The dataset is preprocessed to remove noises like null value, incomplete reviews, and unverified reviews. Then, we built a sentiment analyzer based on DistilBERT to rate beauty products based on the positive and negative scores from the products reviews. Next, we train a skin type model to detect four skin types: dry, oily, combination and natural using a CNN. Then, we trained a recommendation system using a factorization machine to automatically recommend skincare products to users based on the skin types. Lastly, we built a chatbot in Telegram for users to input their facial image for skin detection and product recommendations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
IFTTT	If This Then That
ROI	Regions of Interest
FC	Fully connected
PPG	Photoplethysmography
ReLU	Rectified Linear Unit
CONV	convolutional layer
VGG	Visual Geometry Group
NLP	Natural language processing Net
NPS	promoter Score
CRM	Customer Relation Marketing

CHAPTER 1 – INTRODUCTION

1-1 Problem Statement and Motivation

Various chatbots have developed to recommend beauty skincare products to the customer. This project's **problem domain** is the chatbots that exist now suggest the products to users based on the product description. However, the product description might not be accurate because it is labelled by the manufacturer. The language on the ingredients list on the skincare products uses the International Nomenclature of Cosmetic Ingredients, so the users find it hard to understand what is written. The INCI has existed to create a standardized language of the ingredient names labelled on the product. However, it is not user friendly because some of the manufacturers attempt to appease the users by putting the more familiar name in parentheses next to the scientific name, for example, Alpha-tocopheryl acetate (vitamin E) to stop their complaints. The ingredients list will most likely look like a whole line of foreign language isolated by commas without this. Rather than doing investigator work on the ingredients list, chatbots are here to make our life easier by recommending products to users based on the product description labelled on the products. In this project, a **real-world review is used for chatbots** to recommend products to other related users.

1-2 Project Scope

Nowadays, the customer will look at the product description to determine if the product is suitable. This project develops an AI model to get the customers' actual feedback after using the product so that the users will get the most suitable products for them. When the products are recommended to the user, the AI model does not retrieve information from the product description stated from the product but from the users' reviews. For example, the AI model will not analyse the data of ingredients stated in Aloe Vera Gel by the merchandise and the data of reviews given from the users who have tried the product. The proposed machine learning model is based on user review, and predictions are made based on real reviews according to the user's particular skin type. The machine learning model uses training data that are preprocessed and accurate.

1-3 Project Objectives

The first objective is to **train a skin type classification model using CNN for skin type based product recommendations**. CNN has recently obtained impressive results in a number of recognition activities in the area of computer vision. The CNN structure is built on structures found in the biological brain's visual cortex. The neurons in a CNN are organised in three dimensions (width, height, and depth), and each layer's neurons are only connected to a small portion of the previous layer, the neuron's receptive area. The convolutional layer (CONV) – accompanied by a non-linearity, such as the Rectified Linear Unit (ReLU) –, the pooling layer, and the completely connected (FC) layer are the three basic layers found in most CNNs. The last two fully connected layers of the trained CNN and treated the remaining part of the convolutional neural network as a fixed feature extractor to tune the network for the skin type classification problem. Using the features learned by the CNN, a **linear classifier (softmax) is trained for the skin type classification.**

The second objective is to **train a sentiment analysis model using product reviews data** crawled from cosmetic website. The product reviews are the feedback from the customers after using the skincare product. Customer feedback on the internet is a real treasure to the business that can be used to develop the company and increase customer loyalty. Data is the king of everything in today's digital world. These data are valuable resources for companies in the beauty industry that can help them achieve greater success. Customer reviews on the internet are basically a large database full of useful and valuable information. People also search the internet for feedback based on the experiences of real people before making a purchase. Moreover, reviews are often important in determining whether or not an individual purchases a product or service. Therefore, by designing a system that can suggest beauty skincare products to users based on the real user reviews, it will be very convenient for the consumers and make their life easier.

The last objective is to train a **recommendation system to suggest product types based on user skin types**. Nowadays there are commercials and advertisements everywhere promising the skincare products will make the people prettier and more gorgeous. But the most important thing is to choose the most suitable product for the skin. Different products suite different people. The product that is good for a person may not be good for another. This is because everyone has their very own skin type. The product that is suitable for the particular skin type is recommended to the user. Based on the product reviews from real users, the product which is suitable for the skin type can be determined and recommended to the user using this recommendation system. By inputting the image of the user's portrait into the system, user will know which product is suitable for him/her. Users does not has to waste time assessing their skin type by themselves and do a bunch of survey before purchasing the skin care product.

1-4 Impact, Significance and contribution

The skin care industry is climbing and flourishing day by day and it does not looks like it is slowing down anytime. New devices and treatments are launched onto market daily to give contribution in the beauty industry. The proposed AI modal will bring a huge impact to the beauty skin care industry. When a person wants to purchase skin care products, he/she has two choices: go to the shopping mall and search consultation from the promoter from skin care retailers or stay at home to do research by himself/herself. To purchase the most suitable skin care product, the skin type of the user has to be determined first. In this day and age, there are various ways used to determine the skin type of users. For example, asking them to answer the questions in order to determine the skin type. The user has to observe his/her skin type and determine the complexion. After determining the skin type, the promoter will recommend the products for the user based on their knowledge and experience. This will consume much time of the user to get the right product excluding the possibility of the intention of promoter to recommend expensive products that are not suitable for their very own benefits (sales and commission). User has another option to stay at home and do research on the products that are suitable for their skin type. This will consume a lot of time too and has the possibility of getting the wrong results.

Whereas by inputting the portrait of the user into the proposed AI model, with just a click the user is able to get the information of the recommendation product. It is simple, convenient and brings a lot of benefits to the users. This will save the time of users as they can easily use the recommendation system at home and get the information of product that is suitable for them. The user does not have to answer long questions in order to determine their skin type. All they need to do is just taking a selfie and input the image in the recommendation system, then will have the answer to purchase the skin care products.

1-5 Background Information

There are various beauty skincare products in the market now and it is tough to choose the most suitable product. Since many of the products are practical, skincare technology is diverse and distinguishes itself from many other cosmetic categories. Some products are primarily used for cosmetic purposes and products that are primarily used for therapeutic purposes.

Despite the fact that much of the marketing literature has focused on international brands, there has been very little research on brand loyalty in Asian markets. Against the backdrop of significant shifts in the global brand marketplace, academics seek to understand why customers are loyal to a particular brand and the main factors that affect their brand loyalty. People want a higher standard of living, and advances in technology, science, economics, and education fuel consumer demand for existing styles and new goods.

The general population's income level rises in tandem with the strength of the Malaysian economy. As a result of the increase in wages, a new way of life has emerged. Beauty and skincare products are one product group that has seen a significant rise in demand over the last decade. Malaysians are now more than ever concerned with beauty regimens, such as skin and hair care, out of a desire to look attractive, young, and fashionable. Skincare products are formulated to moisturize, cleanse, tone, and preserve the skin. [1]

In Malaysia, one of the most dynamic product categories in 2014 was skincare. Malaysian cosmetics and skincare products are now gaining popularity. Meanwhile, the younger generation, who have been strongly influenced by Korean music and drama shows, prefer to use branded Korean makeup and skincare items worn by their favourite celebrities. Market attention has been drawn to skincare products due to continuous innovation by industry players, especially foreign brands.

CHAPTER 2 – LITERATURE REVIEW

The previous year has been a sensational expansion in the utilisation and accessibility of beauty chatbots. **Various beauty chatbots developed with different developers' features have been introduced and implemented in the beauty industry** to resolve inaccurate product recommendations. The beauty industries are taking advantage of the numerous and rising benefits of chatbot innovation, and it has been the current practice towards the problem. The beauty industry is one of the numerous industries that has understood this innovation's mind-boggling capability to communicate personally with purchasers, opening another universe of advertising prospects being discovered as the innovation improves over the long run. Beauty is considerably more personal than other industries, and informing succeeds when it is customised — regardless of whether that is personalised product suggestions or beauty tips.

The initial steps of chatbot innovation in the beauty industry were careful and clear. Beauty chatbots were first used to deal with normal client assistance inquiries, such as store locations and item returns. Beauty chatbots have developed into intelligent virtual store aides for many leading fashions, beauty skincare and cosmetic brands. Some even enrol the help of increased reality to offer customised product suggestions and beauty tips and comes with the try-ons.

The best cosmetic beauty chatbots in the business currently effectively handle the role of beauty advisor, appointments, gifting and influencer marketing in major chatbot platforms like Facebook Messenger and Kik. The users can have a discussion with the chatbot similar as they would with a promoter in the cosmetic stores and get tons of beauty ideas, hacks and suggestions. The chatbots ask the users a progression of inquiries, the clients at that point share photographs and are in the progress of connecting to a digital beauty consultation.

2-1 Existing system

Below are the 4 most well-known beauty chatbots currently available to perceive how they contrast in administrations and application. [2]

2-1-1 HelloAva

This famous beauty chatbot goes about as a delight expert by assisting users by tracking down the correct products for **building a customised skincare routine**. The users filled up a survey according to their skin and sent a selfie to the chatbot. HelloAva is available on platforms such as Facebook Messenger and it can even be installed on the desktop computer. HelloAva's calculation suggests a wide assortment of products, and afterwards an authorised aesthetician affirms the decisions before the users reach the checkout page.

Using AI and human expertise, Ava helps find the best products for different skincare needs of the users by giving personalised product suggestions. Users can likewise talk with aestheticians on the off chance that they have any inquiries regarding the suggestions. HelloAva's AI innovation utilises reference information on different customers from different factors such as their photo, demographic groups, current living environment and past product history. It also monitors every user's past conversational history and ongoing buyings to offer persistently personalised support. They are a group of authorised skin care consultants and dermatologists who will assist the users behind the screen of digital applications.

It is clear that this beauty chatbot has given the users a solution for the problem of choosing from the huge range of beauty skincare products. The promoters in the beauty skincare stores will indeed like to promote the products from their store that they are familiar with and to them, the more expensive, the better the products as they will gain high commission after selling them. Using HelloAva, this problem can be solved as it will suggest a suitable product to the users based on the users' different factors and build a customised beauty regimen for them. [3]

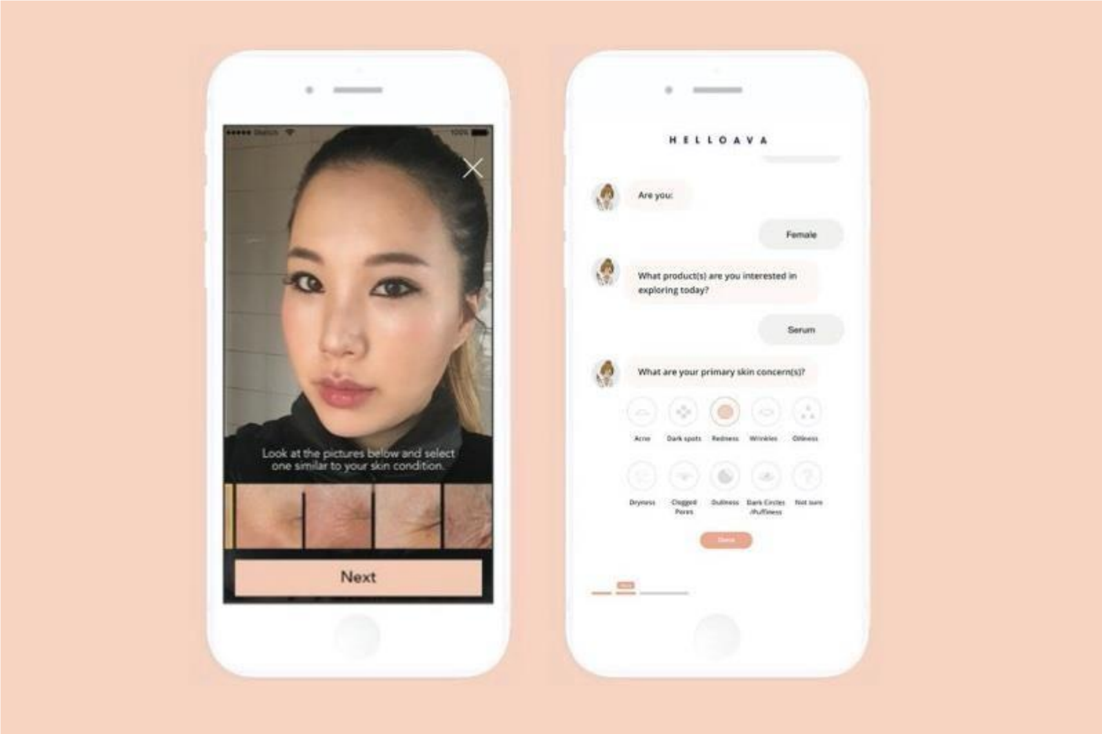


Figure 2-1-1: Screenshot of Chatbot HelloAva

HelloAva is not completely perfect and also have some weaknesses. Firstly, HelloAva works well for these straightforward, well-defined processing scenarios but does now work well for unexpected or uncommon scenarios where there is a breakdown process. Secondly, HelloAva does not have the augmented reality features such as product try-on, looks, virtual tutorials, colour match and swatch me. The product try-on function allows user to try on beauty skincare makeup products. Customers cannot digitally try on the eye, lip, and cheek makeup in a variety of colours and palettes using the Product Try-On function. The lacking Looks feature does not allow the customers to be inspired by and try on different looks produced by experts. The lack of Virtual Tutorials also makes the consumers unable to follow along with experts to learn new strategies step by step. Without the Color Match feature, users can find the ideal makeup colours to match an outfit, a hairstyle, or a personality. HelloAva does not have the Swatch me function, which acts as a virtual arm swatch to try out and compare hundreds of different colours of the skin make up. These limitations can be solved by using the Sephora Virtual Artist chatbot.

2-1-2 Sephora Virtual Artist

Using the Sephora Virtual Artist can let the users get a **virtual makeover** and share their thoughts with their friends and families without any problem. They can **try-on an endless library of eyeshadows, lip tones, and surprisingly fake lashes to track down their ideal shade** and make their lips perfect—and all of these can be accomplished without stepping a foot in a store. Sephora Virtual Artist is a shade matching bot that helps users try out various looks by implementing the filters to select the skin make up products based on category, brand, colour family, formulation or favourites. The users can then save the various looks and view the combination of products applied on the face before purchasing the products. [4]

Bringing computer-generated reality innovation to the excellence business, Sephora's cosmetics application utilises facial recognition for the users to experiment with the skin makeup products anywhere. The application examines the user's face, detects their eyes, lips, and cheeks for the skin makeup products arrangement. Let them try on the makeup virtually to see exactly if this eyeliner or that lipstick resembles the users. For example, if the user has a problem choosing between KVD Vegan Beauty fluid lipstick or Anastasia fluid lipstick, there will be no worries as he/ she can try them both and perceive what they look like before making the decision on which to purchase. This virtual lipstick analyser is popular, especially among women, as it can filter through colours, textures and finishes. This will reduce the problem of customers purchasing the lipstick in the wrong colour.

By trying on the skin makeup products, the users can perceive their full looks virtually by following the step-by-step tutorials. They can customise their face by matching the cosmetics to their outfit and analyse many colour swatches right away. They can avoid guessing work and questioning out of contouring, lip lining, or even the delivery of a smoky eye. The users can observe how the ongoing trend works out on their face and figure out how to reproduce them with clear instructions from Sephora Virtual Assistant. They can learn to blend the different skin makeup products on their face and try on the products to accomplish the look that they favour. With the virtual arm, they can compare colour swatches from various brands and finally choose the one skincare product that is most suitable for them. [5]

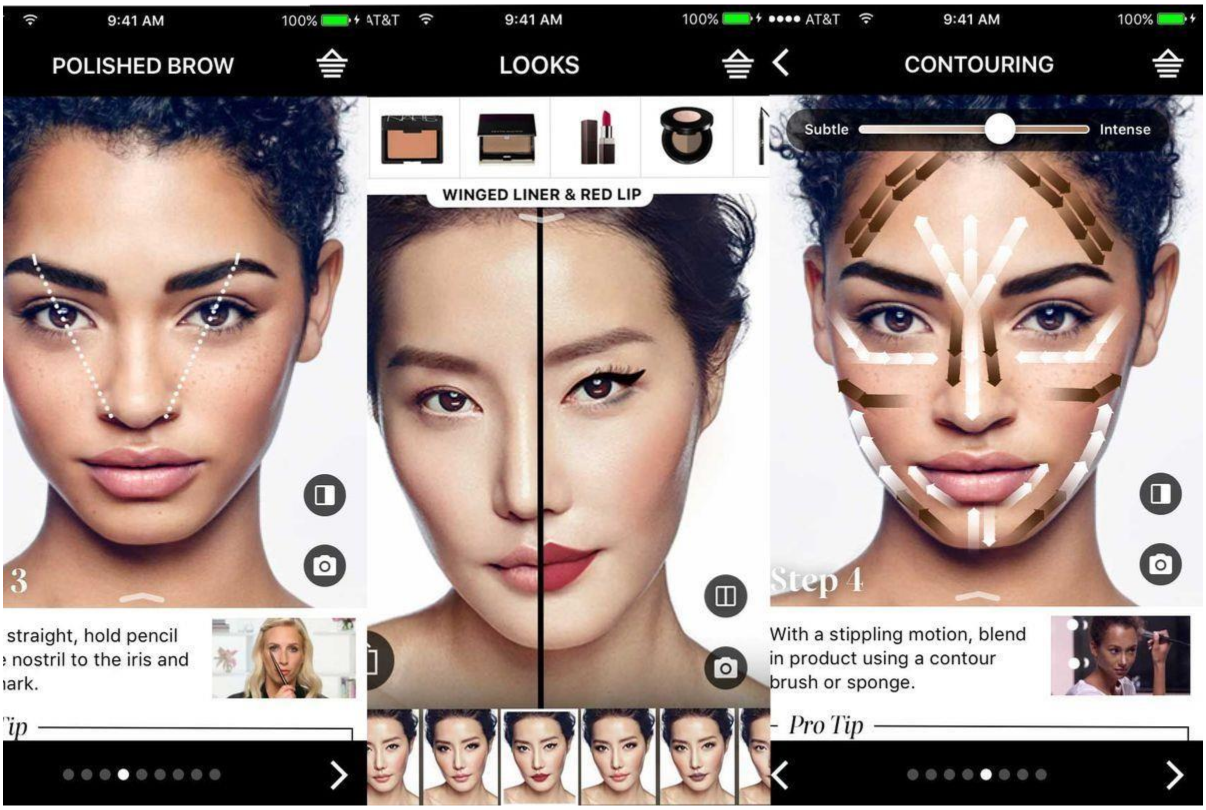


Figure 2-1-2: Screenshot of Chatbot Sephora Virtual Artist

Sephora can recommend skin makeup, but it **cannot recommend skincare products.** Unlike showing the users how their faces look after applying the skin makeup product, users cannot perceive how their face looks after applying skincare products. This is because skincare products are different from skin makeup products that can be observed when applying on the face with naked-eye. Sephora Virtual Assistant can only let users try on different looks with a vast range of makeup shades and tones but does not assist in skincare product recommendation. This limitation can be resolved by using the HelloAva chatbot to build a personalised skincare routine.

2-1-3 CoverGirl's Kalanibot

Kalanibot is the world's first influencer chatbot. Kalanibot is modelled on Instagram influencer Kalani Hilliker, a teenager dancer with more than 4 million followers on Instagram. Up to this point, CoverGirl's business plan has focused on celebrities, especially film stars and fashion models. With the new climb in the prevalence of social media influencers, CoverGirl has created this influencer chatbot.

The chatbot is available through the platform messaging app Kik, and the targeted audience is teenagers. Kalanibot communicates with its users conversationally and promotes the cosmetic products that are used by Kalani. It also offers coupons for cosmetic products that are promoted to the users before they purchase the products. This chatbot helps shoppers browse easily and quickly through the site, and it also helps them in a very personal way. The KalaniBot offers skincare tips and pointers by adding a feeling of the human salesman contact and offering coupons to users that utilise the chatbot.

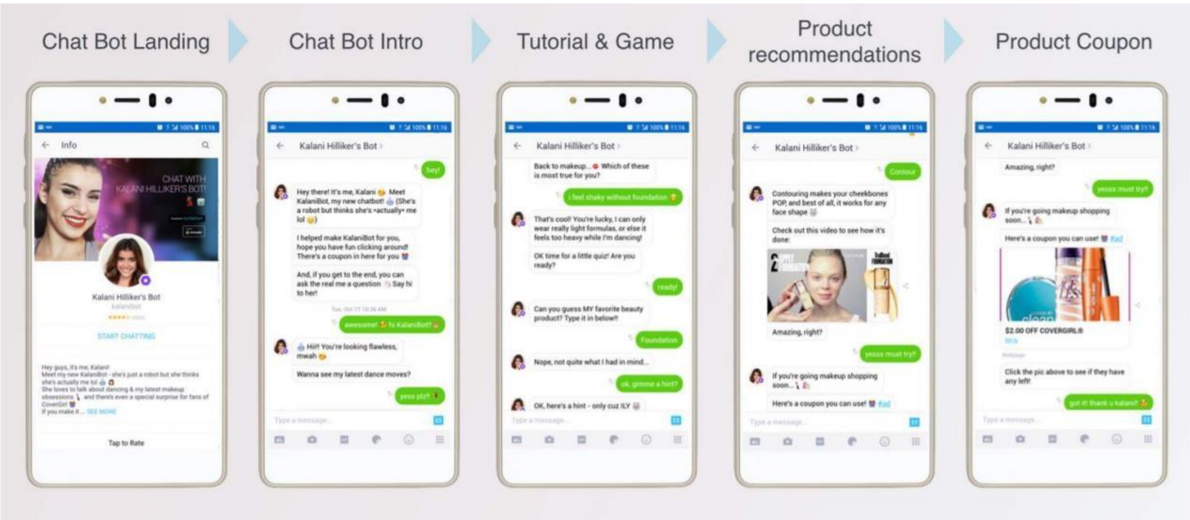


Figure 2-1-3: Screenshot of Chatbot CoverGirl's Kalanibot

The Kalanibot chatbot interacts with the users in the way they want Kalani to and keeps the users at the centre of her focus. In this era where service is mostly neglected as the organisations and brands are only paying attention to income, the creation of Kalanibot brings a refreshing change. This chatbot is not simply consumer-centric, it is also intelligent, entertaining, and idiosyncratic. These features resemble the features of an ideal salesperson in stores. The chatbot mirrors the users' character adored model and artist, Kalani Hilliker, and creates a human connection with the users with its charm and sense of humour.

Both young adults and grown-ups have been crazy over CoverGirl's new chatbot. On top of the extravagant character that pulls consumers towards the site, it has helped the digital magazine earn more money.

With the utilisation of KalaniBot, CoverGirl's click-through coupons reached an achievement rate of 51%! While CoverGirl had depended intensely on email marketing, compared to other ways, it did not acquire the fascinating results that this chatbot did. By following the trend set by KalaniBot, other brands from the beauty industry started utilising the chatbot administration to develop their customer base.

Notwithstanding, the KalaniBot is not 100 percent perfect. The users have a complaint that although the interface and the interaction are wonderful, it does not give exact responses to all questions. Assume a user picks a choice from the given choices provided, KalaniBot might not predict the correct answer using predictive text as the choices are numbered. [6]

2-1-4 L'Oréal's Beauty Gifter

L'Oréal has released a bot called Beauty Gifter that uses Facebook Messenger to sell customised makeup and skincare gifts by asking both the gift giver and the recipient questions.

Beauty Gifter is the first in a series of conversational interactions that L'Oréal hopes to develop with **Automat Technologies**, a Montreal-based startup that has worked largely under the radar since its inception last year.

In a phone interview with VentureBeat, Automat CEO Andy Mauro said that one of the company's upcoming initiatives might be the use of in-store QR codes to facilitate conversational experiences.

The Messenger camera can now check QR codes, thanks to the launch of Messenger Platform 2.0 last month at the Facebook developer conference F8. Because of recent updates to Facebook Messenger, QR codes can now be related to a wider variety of destinations, including unique conversational encounters inside a bot, such as someone searching for a gift rather than someone who wants to go straight to sales. Since L'Oréal owns hundreds of brands, from The Body Shop to Giorgio Armani, Mauro believes that telling the gift recipient what they want would not ruin the surprise.

"Right now, there are a lot of crappy gift bots on the market, and most of them ask the gift recipient to answer certain questions. "Well, the gift giver usually doesn't know the answers to those questions, but the gift recipient usually does," he said. "A lot of people say, 'Hey, it'll ruin the surprise,' but we've noticed that the majority of people don't mind. It doesn't really ruin the surprise because there are a dozen L'Oréal brands involved."

L'Oréal Beauty Gifter provides one-on-one talks with users at scale and launched its chatbot through the platform Facebook Messenger. The chatbot asks the users a progression of inquiries, such as the **price range of the gift** and the **age of the person who will receive the gift**. After that, a message that states, "I want to send you a gift." will be sent to the receiver of the gift. [7]

The bot will then asks the recipient a progression of inquiries concerning their skin tone and type, what colour combinations they favour, etc. When the inquiries are made, the bot sends a progression of gift choices to the gift giver. Through AI, Beauty Gifter learns every user's inclinations and makes customised product suggestions from the world's leading beauty and cosmetic brands. L'Oréal Beauty Gifter follows up with personalised, relevant content to create strong commitment with the users. [8]

This chatbot's weakness is it mainly just provides the service of gifting the skincare products while lacking focus on the beauty product recommendation. There is user input automation, but it is simple and based on the choices chosen by the user to choose the product. The product recommendation inclines more to the user favorable products chosen by the user due to lack of facial recognition system. The user has to identify by himself/herself the skin toneand type, and it is not identified by the chatbot itself. This limitation can be solved by implementing the facial recognition feature.

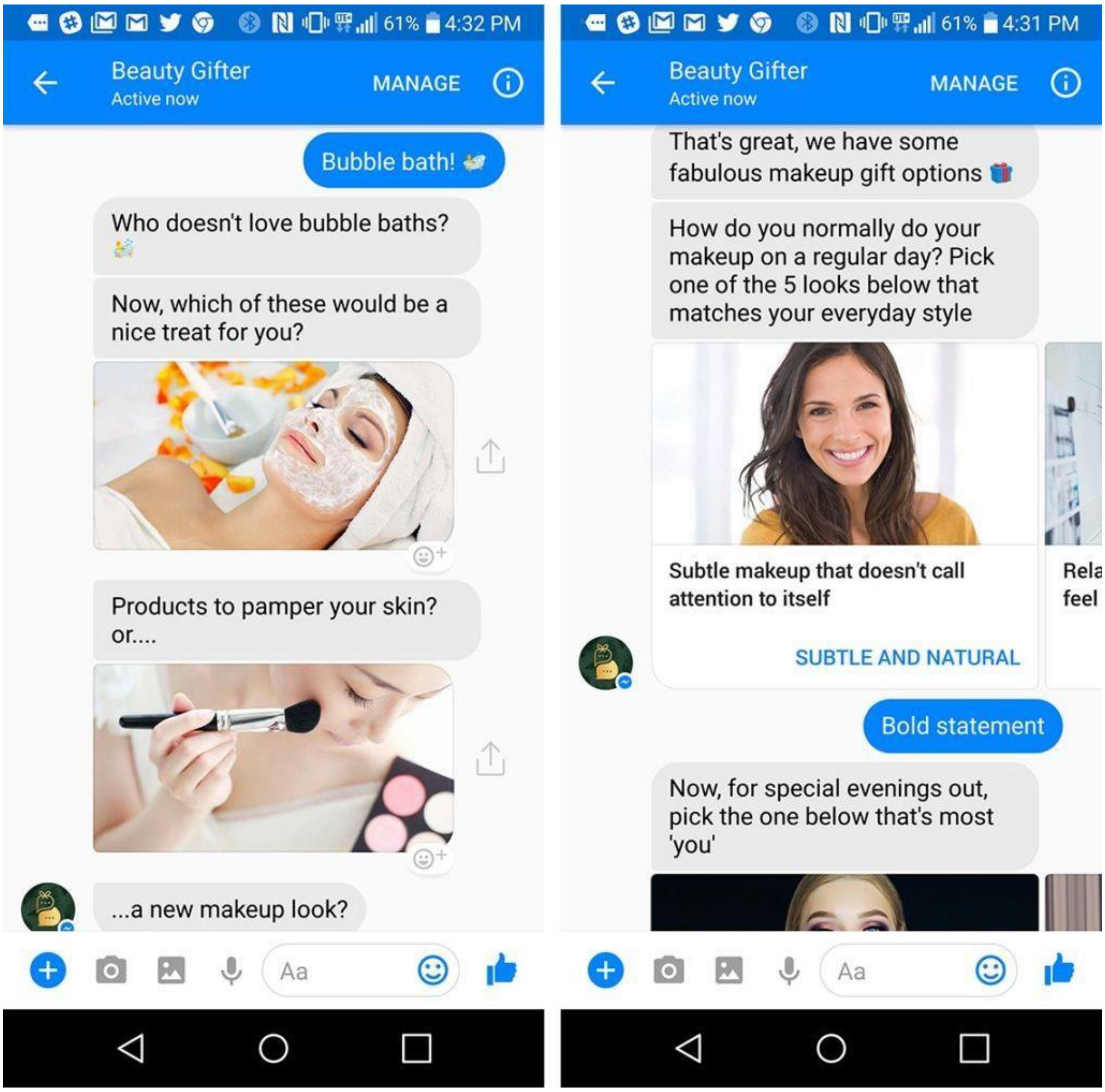


Figure 2-1-4: Screenshot of Chatbot L'Oréal's Beauty Gifter

2-2 Chatbots Comparison

	HelloAva	Sephora Virtual Artist	CoverGirl's KalaniBot	L'Oréal Beauty Gifter	Proposed Chatbot
Personalised skincare product recommendation	✓		✓	✓	✓
Skincare product recommendation based on real user reviews					✓
User input automation	✓	✓	✓	✓	✓
Facial recognition	✓	✓			✓
Classifies user's skin type	✓				✓
Range of product	Skincare	Skin makeup	Skincare /Skin makeup	Skincare /Skin makeup	Skincare /Skin makeup
Try-on makeup products		✓			
Gifting				✓	

Table 2-2: Beauty Chatbots Comparison

HelloAva assists user by tracking down the correct products for building a customised skincare routine. As compared to the proposed bot which uses real user reviews for product recommendation, in HelloAva user needs to fill up a survey according to their skin. The proposed chatbot is more convenient and simple to use comparing to HelloAva as user just needs to send a selfie to chatbot and knows their skintype immediately without filling up a bunch of questions. HelloAva does not have the augmented reality features such as product try- on, looks, virtual tutorials, colour match and swatch me. These limitations can be solved by using the Sephora Virtual Artist chatbot. The weekness of Sephora Virtual Artist is it it cannot recommend skincare products whereas by using the proposed chatbot, user can get their very own skincare product recommendation. . The KalaniBot offers skincare tips and pointers by adding a feeling of the human salesman contact and offering coupons to users that utilise the chatbot. It emphasize more on pleasant interaction with users and does not give exact responses to all questions. The proposed chatbot has a clearer and straightforward aim to recommend user the beauty skin care products as compared to KalaniBot. L'Oréal's Beauty Gifter product recommendation inclines more to the user favorable products chosen by the user due to lack of facial recognition system. The user has to identify by himself/herself the skin tone and type in L'Oréal's Beauty Gifter which is not user friendly as compared to the proposed chatbot that can identify itself the skin type of user and make product recommendations.

CHAPTER 3 – SYSTEM METHODOLOGY

3-1 Methodology

A methodology is proposed to realize the project in the projected timeframe.

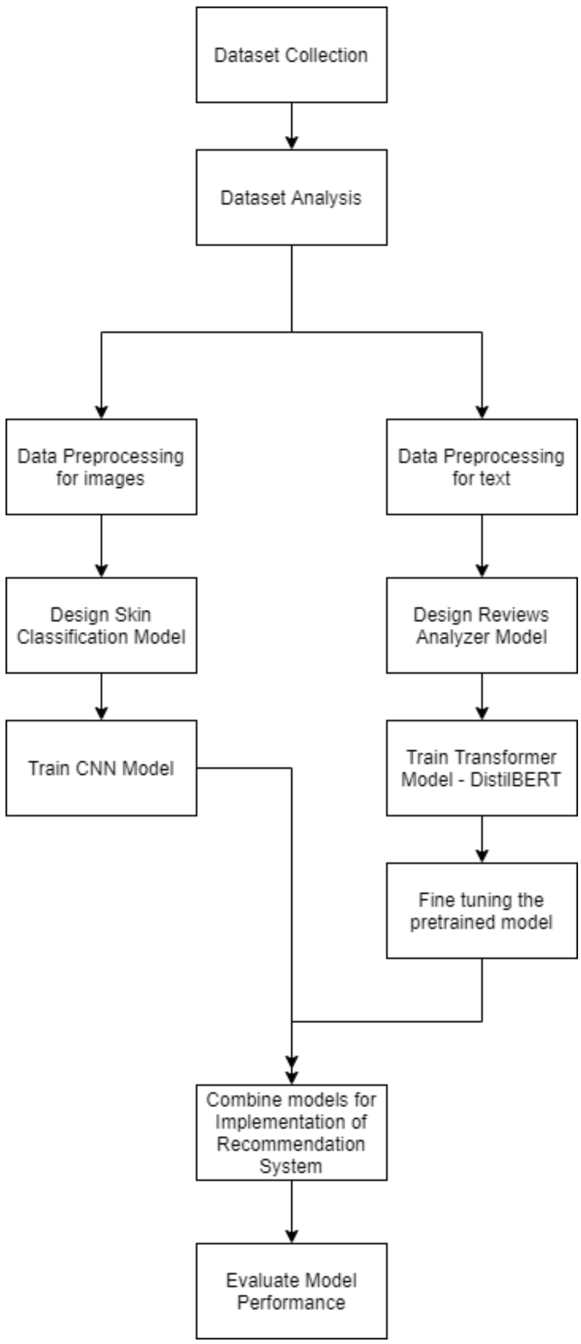


Figure 3-1: Methodology

3-2 Datasets

3-2-1 Image Dataset for CNN Model

Multiple datasets are downloaded from the website Kaggle.com. Datasets are also collected manually from various websites to increase the number of data. The datasets for the image are classified manually according to the skin type. For the data preprocessing of image, the images are resized to a consistent size and BGR are converted to RGB colour scheme. This is because when converted to RGB, it will be saved as a correct image even if it is saved after being converted to PIL. Image object. Several image processing libraries have different pixel orderings, therefore the conversion is needed to ensure the consistency of the dataset.

Figure 3-2-1 : Dry skin image

Dry – Dull and dehydrated skin, skin is less elastic and has rough complexion

Figure 3-2-2 : Oily skin image

Oily – Skin with large pores and shiny with thick complexion. Pimples and other blemishes appear on skin.

Figure 3-2-3 : Combination skin image

Combination – Tends to be oilier in the T-zone (forehead nose, chin), while drier on the cheeks

Figure 3-2-4 : Natural skin image

Natural - Skin is healthy and well hydrated, barely visible pores

3-3 Text Dataset for Sentiment Analysis

product_id	review_title	review_text	rating	age_range	skin_type	skin_tone	eye_color	reviewer_username	tags	review_id
P38217	Worth the m	Sometimes I stray from	5		normal	light		katechatte	{foamy,exfoliating	6611717f-2636-4756-bf36-66c81cc267a7
P38217	Great	I am a 41 year old Africa	5		combinati	deep		snook41	{foamy}	e7d3307e-02ff-45a1-8fc3-6bd628bedd86
P38217	Great Produc	I'm really enjoying this	5		combinati	olive		wahinewarrior	{foamy,milky,exfol	4188d728-fde6-4d06-984e-164cca2b8781
P38217	Nice, but not	I tried this cleanser at a	3		combinati	fair		jenlines22	{hydrating,creamy	248c904c-6e30-4929-8228-87b03ad7a921
P38217	great moistu	leaves the skin feeling f	5		dry	light		jessea	{exfoliating}	654bdb99-9371-4440-a540-0dd2a73da339
P38217	It works	I am a 33 year old Latin	5		combinati	medium		Anonymous	{milky}	35eadf21-f589-4e57-b149-af59a2e8fe07
P38217	Feels great!	My best friend sent this	5		combinati	light		wahine79	{milky,exfoliating,	938f2f6e-fcb6-4a51-95ef-a55374412a74
P38217		very good cleaning-rich	5		combinati	medium		cheeta	{foamy,milky}	bd51ed5c-cdec-4f42-873b-e07f4b79718d
P38217	Great Face W	Love it! Gentle and non	5		combinati	light		lvthesun	{foamy}	d43691fb-9799-4c0e-90fd-8fdd03900ade
P38217		i like it, but not so crazy	3		combinati	medium		tweezerama	{foamy}	72c2a20a-be1c-46dc-a2cd-142f2110a05d
P38217	Pleasant Sup	I'm a 42 year old Africar	5		combinati	deep		kathylove	{foamy,milky,exfol	218fcf01-2a13-47c4-b3ab-179dd1edd0a3
P38217	very suprisec	it left a tight and clean f	4		normal	light		chikostik		4c2753dc-65bb-4839-baab-cb2716f0890c
P38217	love this	I love this cleansing cre	4		combinati	medium		tequilayoko	{foamy,exfoliating,	4042d55a-8737-45c5-9ea8-bbb8bb174a15
P38217	good	nice smell, texture, and	4		dry	light		superjunior	{milky,hydrating}	a81de143-0b68-467a-b865-b5cbb3a619f8
P38217	A little bit go	I have many cleanser (fi	4		combinati	medium		akane	{foamy}	4915e064-af66-46ba-99d0-bea7343b40c3
P38217	Finally!!	A non-drying cleansing	5		combinati	light		njsephoramom	{exfoliating,cream	deecd6ce-8dfc-4128-9eaf-8837264a3c5e
P38217	Does what it	This cleanses well and c	4		combinati	light		mlparr	{exfoliating,milky}	432f0fca-fccc-4a1f-aa16-2b44ce2db276
P38217	smooth face	I love this cleansing cre	5		combinati	light		texasbeach	{foamy,milky,exfol	51f65a53-66e7-4c63-be5a-ff906464936c
P38217	Stings my eye	Returned this product a	1		normal	light		nancihi	{foamy}	ebf96058-5203-4494-9a6c-2bb74099ad23
P38217	Great Cleans	This is a great cleanser.	5		combinati	fair		carpemakeup	{foamy,exfoliating	cdb9eec2-143b-440a-a3e1-ea6f521e8250
P38217	fantastic clea	I love this cleanser by M	5		dry	medium		classicbeauty1	{foamy,"also---low	655b16ad-0250-4917-8465-bb6e068a61f1

sephora_review Product 1 Product 2 Product 3 Product 4 Product 5

Figure 3-3-1 : Text Dataset Reviews from Actual Users

The are 1000 datasets with the attributes product_id, review_title, review_text, rating, age_range, skin_type, skin_tone, eye_color, reviewer_username, tags and review_id. For the 'skin_type' attribute, it is divided to 4 skin types which are oily, dry, combination and natural. For the ‘rating’ attribute, there are rating valuee 1 to 5, an assumption is made where rating 1,2 are negative reviews and 3,4,5 are positive reviews. Data is preprocessed by dividing the data to new excel sheets according to their unique product_id. The null values in the data are manually removed. Data is cleaned and data augmentation is applied as part of data preprocessing. Tokenization of data for input to the distilBERT model is done.

id	product_url	sku	category	brand	name	rating	detail_text	size_oz	price
P0249	/single-eye-shadow-F	221135	eyeshadow	NARS	Single Eye Shadow	4.3354	What it is: A collectio	0.07	26
P0250	/duo-eyeshadow-P02	221309	eyeshadow	NARS	Duo Eyeshadow	4.4772	What it is:A mini min	0.14	36
P0277	/liquid-eyeliner-P027	91629	eyeliner	Dior	Liquid Eyeliner	4.1841	An automatic, total-c	0.04	35
P0417	/double-ended-blemi	1408400	blemish & acne treat	SEPHORA COLLECT	Double-Ended Blemi	3.5861	What it is:A long-handled, stainless steel, c		17
P0766	/l-heure-bleue-P0766	216614	perfume	Guerlain	l'Heure Bleue	4.38	The breathtaking bea	1.7	83
P0771	/shalimar-eau-de-toil	556076	perfume	Guerlain	Shalimar Eau de Toi	4.6042	Guerlain's most succ	1.7	82
P0782	/amarige-P0782	6494	perfume	Givenchy	Amarige	4.4643	Fragrance Family: Fl	1	56
P0845	/pour-femme-P0845	78600	perfume	BVLGARI	Pour Femme	4.6753	In 1994 the renowne	3.4	140
P0847	/eau-parfumee-au-th	7237	perfume	BVLGARI	Eau Parfumée Au Th	4.4741	BVLGARI's first colog	1.33	72
P0901	/pour-homme-P0901	13375	cologne	DOLCE&GABBANA	Pour Homme	4.6574	Introduced in 1994,	2.5	67
P0950	/givenchy-gentleman	9589	cologne	Givenchy	Givenchy Gentleman	4.7857	In 1975, Hubert de G	3.3	89
P0956	/pi-P0956	288084	cologne	Givenchy	Pi	4.6792	Pi is a celebration of	1.7	65
P0960	/xeryus-rouge-P0960	9795	cologne	Givenchy	Xeryus Rouge	4.875	In 1996, Givenchy c	3.4	89
P0974	/hanae-mori-butterfly	287532	perfume	Hanae Mori	Hanae Mori Butterfly	4.6793	Hanae Mori, the grea	1.7	83
P0999	/pour-homme-P0999	627711	cologne	Kenzo	Pour Homme	4.5	The first morning of t	3.4	86
P100101	/crushed-cabernet-sc	1257419	scrub & exfoliants	Caudalie	Crushed Cabernet Sc	4.4863	Which skin type is it	5.3	38
P101916	/firm-tone-serum-P10	844357	cellulite & stretch ma	Murad	Firm and Tone Serum	3.6533	What it is:A breakthro	6.75	60
P101917	/firm-tone-dietary-sup	844365	cellulite & stretch ma	Murad	Firm and Tone Dieta	3.9286	What it is:A patented, clinically proven int		138
P102503	/wrinkle-revenge-resc	844480	moisturizers	DERMAdoctor	Wrinkle Revenge Res	4.2366	What it is:A lightweig	1.7	57
P102504	/wrinkle-revenge-eye-	844472	eye creams & treatm	DERMAdoctor	Wrinkle Revenge Eye	4.0591	What it is:A super-hy	0.5	50
P102813	/hypnose-waterproof-	834028	mascara	Lancôme	HYPNOSE WATERP	3.9614	Control the waterproof volume up to 6 time		28
P104006	/blush-bronzer-duo-P	832642	bronzer	NARS	Blush/Bronzer Duo	4.4887	What it is: An all-in-o	0.35	42
P104340	/moisturizing-lip-balm	1696046	lip balms & treatmen	L'Occitane	Moisturizing Lip Balm	3.6498	What it is:An ultra-ric	0.4	12
P104914	/diorskin-airflash-spra	1851013	foundation	Dior	Diorskin Airflash Spra	4.4027	What it is: An ultralig	2.3	62
P107306	/renewing-eye-cream	769836	eye creams & treatm	Murad	Renewing Eye Cream	4.0706	What it is: A multiact	0.5	80
P107319	/cleansing-softening-	763706	body wash & shower	L'Occitane	Cleansing And Softe	4.4493	What it is:A decaden	8.4	25

Figure 3-3-2 : Product Details Dataset

The details of each product can be searched using the unique product_id in this dataset. Each product_id is unique and represents a unique skin care product.

3-4-1 Cross-Entropy Loss

The formula to calculate the cross-entrophy loss.

$$CE = -\sum_{i}^{C} t_i log(s_i)$$

Figure 3-4-1-1: Cross-entropy Loss Function

The t_i and s_i are the groundtruth and CNN score for each of the classes I in C. The activation function Sigmoid or Softmax is applied to scores before performing the CE loss computation. The $f(s_i)$ is written to refer the activations. The binary classification problem $C' = 2$,the Cross Entropy Loss can be evaluated as the figure shown below:

$$CE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1 - t_1) log(1 - s_1)$$

Figure 3-4-1-2 : Cross-entrophy Loss Function in binary classification problem

In the project, we have divided the skin type classification into four classes which are dry skin, oily skin, combination skin, and natural skin. These classes are labelled as C_1, C_2, C_3, C_4 respectively. The groundtruth and score are T1 [0,1] while the groundtruth and score for C_2 is $t_2 = 1 - t_1$ and $S_2 = 1 - s_1$. The Multi-Label classification problem is split in C binary classification problems.

3-4-2 Categorical Cross-Entropy Loss

Softmax is a function that squashes a vector in range of (0,1) and the resulting elements are added up to 1. It is applied to output scores s. The elements act as a class and can be explicated as class probabilities. As the softmax function depends on all elements of s, it cannot be applied independently to each S_i. For the given class S_i, the function of Softmax is calculated as the figure below.

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$$

Figure 3-4-2-1: Softmax Function

Each class in C infer the scores of S_j, the softmax activation for a class S_i depends on the scores in s. The activation functions are used to convert vectors before the computation of loss in the training phase. During the testing of model, the loss is no longer applied. Besides that the activation functions are used to get CNN outputs too.

A loss function used in multi-class classification tasks is called categorical crossentrophy. It is also called Softmax activation with a Cross-Entropy loss. By using this loss, a CNN model is trained to output a probability over the C classes for each of the image. It is used for the classification of multi-classes.

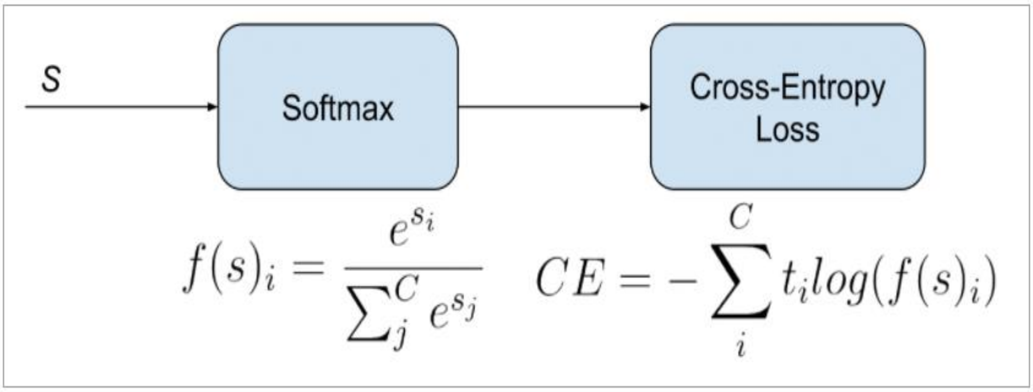


Figure 3-4-2-2 : Categorial Cross-Entropy loss function

The labels in the multi-class classification are one-hot, therefore only the C_p with a positive class can keep its term in loss. Only one component of the Target vector t which does not equals to zero $t_i = t_p$. The elements of summation are discarded which equals to zero cause by the target labels can be written as the function below where S_p is the score of CNN for positive class.

$$CE = -log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right)$$

The skin images datasets are input and used to train the CNN model. The example of dataset can only belong to one of the categories and the trained model must decide which category it belongs to. [9]

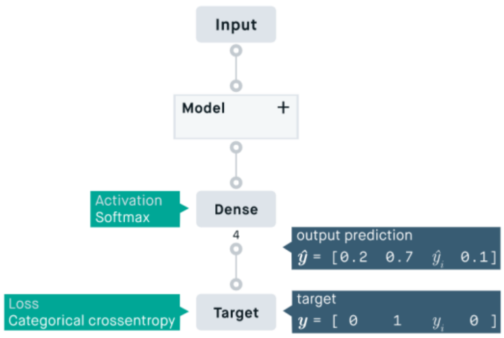


Figure 3-4-2-3 : The flow of Categorical Crossentropy Loss

After defining the loss, the gradient respect to the output neurons of the CNN model is computed to originate it through the net and optimize the defined loss function that is tunes the net parameters. The gradient of CE Loss that is respected to each of the CNN class score in s needs to be evaluated. The loss terms that come from negative classes equals to zero. Due to the Softmax of positive class is dependent to the negative classes scores, the loss gradient respected to the negative classes is not cancelled. The gradient expression is same for all C with the exception of ground truth class C_p due to the score of C_p (s_p) is in nominator. The derivative respect to positive class is shown below after the computation.

$$\frac{\partial}{\partial s_p}\left(-log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right)\right) = \left(\frac{e^{s_p}}{\sum_j^C e^{s_j}} - 1\right)$$

The formula for derivative respected to the negative classes are shown below:

$$\frac{\partial}{\partial s_n}\left(-log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right)\right) = \left(\frac{e^{s_n}}{\sum_j^C e^{s_j}}\right)$$

3-4-3 Binary Cross-Entropy Loss

The Binary Cross-Entropy Loss which is also named as Sigmoid Cross-Entrophy Loss is used in this project for sentiment analysis. It is Sigmoid activation with the Cross-Entrophy loss unlike the Softmax loss which is independent for each of the vector element or called class. This means that the loss calculated for each of the CNN output class is not affected by the other element values. The multi-label classification is also used which means that the element that belongs to the class does not affect the decision of classification of the other element that belongs to other classes. A binary classification problem is set up between $C'=2$ classes for each of the class in C. The formula for binary cross entropy loss is shown below. [10]

$$CE = -\sum_{i=1}^{C'=2} t_i log(f(s_i)) = -t_1 log(f(s_1)) - (1-t_1) log(1-f(s_1))$$

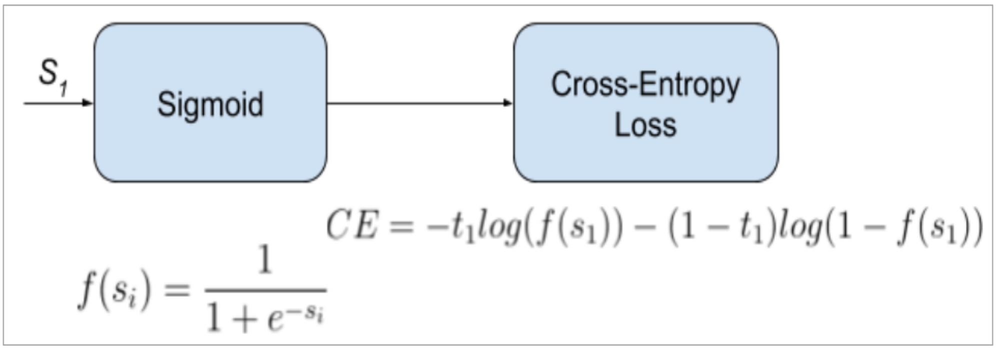


Figure 3-4-3-1 : Sigmoid Cross-Entropy loss

The pipeline for each of the C classes is shown above. The C independent binary classification problems is set to (C′=2). The loss is summed up over the distinctive binary problems. The gradients for every binary problem is summed up to backpropagate and originate, the losses used to monitor the global loss. The class C_1 or is called C_i in C have the score s1 and groundtruth labelled t_1. The score and groundtruth label of class C_2 is $t_2 = 1 - t_1$ and $S_2 = 1 - s_1$.
The loss is evaluated as below where $t_1 = 1$ and class $C_1 = C_i$.

$$CE = \begin{cases} -log(f(s_1)) & if \quad t_1 = 1 \\ -log(1-f(s_1)) & if \quad t_1 = 0 \end{cases}$$

The activation function is dependent in scores of classes in C more than $C_1 = C_i$. The gradient respected to each of the score $s_i = s_1$ is evaluated as shown below:

$$\frac{\partial}{\partial s_i}(CE(f(s_i)) = t_1(f(s_1) - 1) + (1-t_1) f(s_1)$$

The f() is sigmoid function:

$$\frac{\partial}{\partial s_i}(CE(f(s_i)) = \begin{Bmatrix} f(s_i) - 1 & if \quad t_i = 1 \\ f(s_i) & if \quad t_i = 0 \end{Bmatrix}$$

3-5 DistilBERT model

The DistilBERT model is a Transformer model trained by distilling BERT base. It is smaller, cheaper, faster and lighter. It has less parameters as compared to bert-base-uncase by 40%, runs faster by 60% and preserve the performance more than 95% as compared to BERT's evaluated with the GLUE language understanding benchmark. In this project, a smaller language representation model which is DistilBERT is pretrained and fine-tuned with better performances. To utilize the inductive biases using larger models when pretraining, the triple loss combining language modeling, distillation and cosine-distance losses are introduced.

3-6 Text Data Augmentation with back translation

The technique data augmentation is used by creating extra slightly modified version of existing data. The text data augmentation back translation in NLP is performed by giving the input text in a source language. For example, English. Then the text is translated to a temporary location language, for example, the language English is converted to French. Lastly, the previously translated text is translated back to the source language.In this case French is translated back to English. The hugging Face's transformers library is needed to be implemented to perform text data augmentation with back translation. This will increase the number of datasets and thus increase the accuracy of the model and improves the model performance.

3-7 Telegram Chatbot

Telegram bot is a chat messenger service that has the ability to transmit texts, photos, videos, audio files, location information, contacts and documents. It is cloud-based and prioritises security and speed. Telegram has grown in importance as a medium for businesses to focus on. Social media platforms are increasingly being used by organisations to deliver customer support services (CSS). Therefore Telegram has become an important platform to focus on as its user base has grown.

From time to time, we all hear about data breaches involving practically in every social networking application. Telegram bots occasionally collect data from users in order to offer the most appropriate response possible, and when personal data is involved, many people become nervous. Telegram, on the other hand, is a very secure programme due to the added benefits of data encryption, security, and privacy. The same is true for Telegram bots, which means that every message exchange between the user and the Telegram bot is encrypted end-to-end. Only the sender and receiver have access to the messages when they use the pair-to-pair security protocol. These two are, in the instance of the telegram bot, the user and the telegram bot itself.

Telegram is considered fairly secure in light of recent worldwide trends about hacking cases on every social media platform, owing to the fact that messages are sent across the platform in encrypted form. This is a compelling argument to utilize it for commercial purposes, as it will protect all of the customers data. It will provide a safe and secure environment for the users to engage about products and services.

In the proposed project, the users have to send their own facial images to the chatbot to identify their skin type and beauty skin care product recommendation. With the security of telegram bot, the facial images of the user is encrypted and the privacy of the user is protected. Therefore, telegram chatbot is chosen as a platform for the skincare recommendation system. [11]

3-8 RiveScript

RiveScript is a text-based scripting language used in the development of interactive chatbots. It uses plain text, line-based scripting language for simple learning, so that the codes can be typed quicky and easily read and maintain. As compared to other chatbot languages that uses ugly XML codes such as AIML language and codes that require the of memorization of large number of random symbols and “line noises” such as ChatScript, RiveScript is a more simple and user-friendly scripting language for Chatbots. It is chosen to be used in the proposed project as it contains many benefits. It is simple, powerful, flexible, and has the availability of open source. RiveScript exposes a simple plain text scripting language that is simple and easy to be learn and begin with. It is powerful as it has various simple rules that can be merged in powerful ways to develop an amazing chatbot. It is flexible as it takes the “Unix-like” approach for the development. It takes the input of human and replies with an intelligent response. The open source of RiveScript is available for various interfaces such as Java, JavaScript, Python and so on and released under the liberal open source license which is free for anyone to modify, redistribute and use it. [12]

3-9 Tools to Use

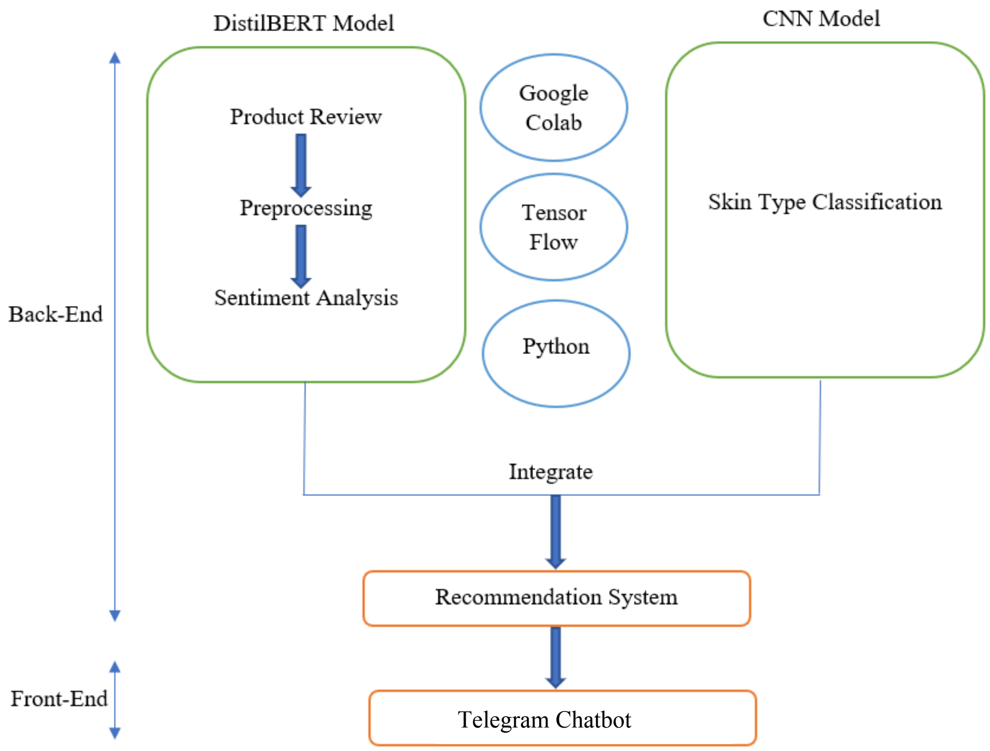


Figure 3-9 : Overall Methods/Technologies Involved

For the back-end development, the model is trained using **Google Colab** which is a free cloud service that supports free GPU. It is a web-based IDE for python and it enables Machine Learning with storage on cloud. It is released in 2017 and brings huge impact to the machine learning and AI world. Popular libraries such as Keras and TensorFlow are used when developing this machine learning model on Google Colab. [13] The machine learning framework **Tensorflow** is usedto ease the process of acquiring the users' review data, training models, serving predictions and refining the future results. In this proposed project, TensorFlow is used to provide a collection of workflows for developing and training models using the language **Python**, as well as deploying them in the cloud, on-premise, in the browser, or on mobile devices. TensorFlow is a free and easy-to-use numerical library for anyone who needs it. It was developed and made available to the public under the Apache 2.0 open source license and by Google. TensorFlow is an open-source machine learning framework at the end of the line. It has a highly inclusive ecosystem of software, libraries, along with powerful development environments, for researchers to keep up with the bleeding-edge as well as developers to design and deploy machine learning- powered applications. TensorFlow was created by those inside Google's Artificial Intelligence department researching machine learning and deep neural networks in the Machine Learning group at Google to furtherexpand on the initial research done to the Google Brain. The system is general enough to be used in other domains, as well. Stable Python and C++ APIs are provided by TensorFlow. [14]

A skin type classification model is trained using CNN for skin type based product recommendations. For modal frequency extraction, a novel **CNN-LSTM based computer vision** architecture has been created. When it comes to processing vibration videos and extracting modal frequencies, the architecture described here is completely self-contained. CNN-LSTM (Convolutional Neural Network, Long Short-Term Memory) deep learning-based approach serves as the backbone for computer vision-based vibration measurement techniques. The use of CNN and LSTM modules resulted in an improved ability to take advantage of Spatio-temporal knowledge that is found in the images. By working with each pixel's small amount of stored information, a more accurate model is obtained for detecting the number of frequencies and reduce the sophistication that is needed to identify them. Spatial information should be represented using each sensor's allocated by way of finding the modal frequencies in a source image, as opposed to each source image finding a separate spatial frequency within the image. Non-contacting, low-cost, high-re, and high-rez zoomable position sensor does not affect the device dynamics, which is a benefit in and is their own right as it doesn't physically interfere with the dynamics in the computer vision-based machine learning model. There is evidence to show that "pixel- sensor" works better, is less error-prone, and is much more effective. Several structures have been used to validate the robustness of the model, as well as dimension size and performance, showing that it can handle a wide range of materials with great effectiveness. [15] An action recognition system, as an example of which would use CNN at the video layer and give each frame a separate representation using LSTM network units to operate on frame-level features is suggested, which then they combine to achieve piece-level features. Also, features are accessed that could arise from the components' movement, which includes both the frequencies and amplitudes of structural videos could uncover. The machine learning-based approach is more accurate through an application of computer vision-enhanced deep processing in an entire system contextually meaningful manner. Basically, this deep learning model is used to replace all image and video processing procedures to directly decompose the structure vibration camera image into the basic vibration modes.

For the front-end development, **Python language** is used to create a **telegram bot**. A telegram app is needed to be installed and a telegram account is needed to be created using mobile number. The python-telegram-bot is needed to be installed as the Telegram Bot API is accessed by this library, which is written entirely in Python. Python versions 3.6.8 and up are supported. This library includes a variety of high-level classes in addition to the pure API implementation, making bot programming simple and uncomplicated. The "telegram.ext" submodule contains these classes. **Botfather** is used to create new bot accounts and managing the existing bots. The Beauty Skin Care Product Chatbot is easily created with the botfather. **Botfather** is a general-purpose automation framework created by a group of European students. Botfather was originally established as a way for computer science students to compete in designing bots for casual games. It has since been expanded to include Android, Browser, and Desktop app automation. Botfather can be used to create automated tests for websites, apps, and desktop applications. Scripts can be executed using Botfather. They include instructions for controlling Android devices and emulators, as well as browsers and desktop applications. Scripts are written in JavaScript and use botfather's functionalities. The language **RiveScript** is used in google collabs for the telegram chatbot. RiveScript allows the interaction of the user with the telegram chatbot.

3-10 Gannt Chart

WBS NUMBER	TASK TITLE	START	ETA	DURATION	Project 2										
					24/1	31/1	7/2	14/2	21/2	28/2	7/3	14/3	21/3	28/3	4/4
	FYP 2														
1	Enhance Model														
1.1	Collect Datasets	24/1	30/1	7 days											
1.2	Retrain and Retest Model	31/1	20/2	21days											
1.3	Performance Evaluation	21/2	27/2	7 days											
2	Chatbot														
2.1	Develop Chatbot	28/2	13/3	14days											
2.2	Testing Chatbot	14/3	20/3	7days											
3	Integration														
3.1	Combine backend and frontend developments	21/3	27/3	7days											
3.2	A/B Testing	28/3	3/4	7days											
4	Report Writing														
4.1	FYP 2 Report	4/4	10/4	7 days											

Table 3-10 : Gantt Chart of Project 2

CHAPTER 4 – SYSTEM IMPLEMENTATION

4-1 Image Classification for CNN Model

Data validation is made where 90% of the data are used to train the model and 20% for validation of data.

```
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=2)
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

```
(397, 128, 128, 3)
(397, 4)
(100, 128, 128, 3)
(100, 4)
```

The images of the dataset are labelled where:

1.0.0.0 → combination skin type

0.1.0.0 → dry skin type

0.0.1.0 → oily skin type

0.0.0.1 → natural skin type

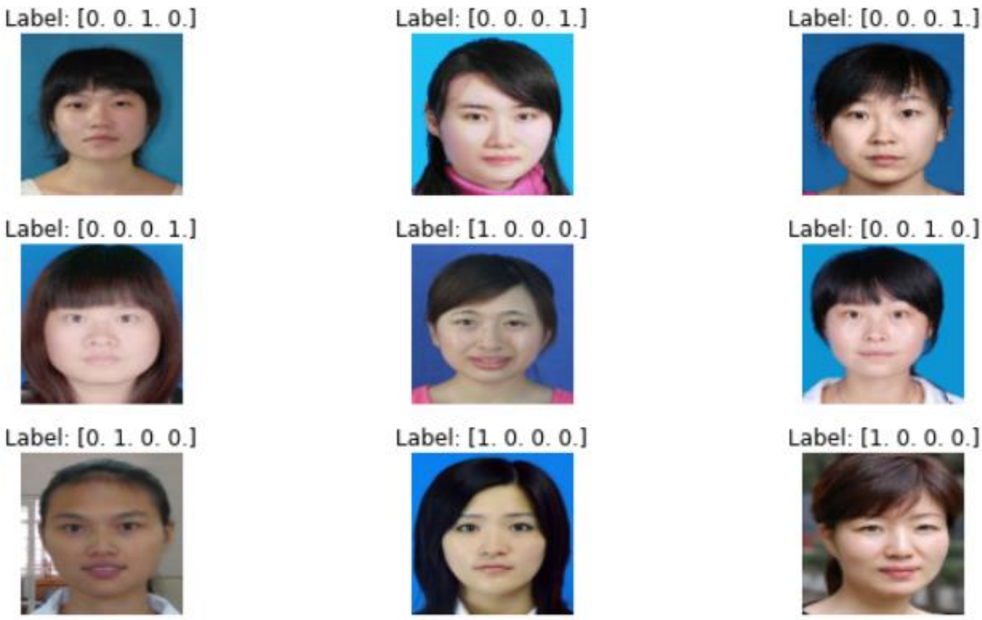


Figure 4-1-1 : Data labelling

The output results of the trained CNN model.

```
Model: "sequential_23"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_176 (Conv2D)         (None, 128, 128, 32)      896

 conv2d_177 (Conv2D)         (None, 128, 128, 32)      9248

 max_pooling2d_66 (MaxPoolin  (None, 64, 64, 32)       0
 g2D)

 conv2d_178 (Conv2D)         (None, 64, 64, 64)        18496

 conv2d_179 (Conv2D)         (None, 64, 64, 64)        36928

 max_pooling2d_67 (MaxPoolin  (None, 32, 32, 64)       0
 g2D)

 conv2d_180 (Conv2D)         (None, 32, 32, 128)       73856

 conv2d_181 (Conv2D)         (None, 32, 32, 128)       147584

 max_pooling2d_68 (MaxPoolin  (None, 16, 16, 128)      0
 g2D)

 conv2d_182 (Conv2D)         (None, 16, 16, 256)       295168

 conv2d_183 (Conv2D)         (None, 16, 16, 256)       590080

 flatten_23 (Flatten)        (None, 65536)             0

 dense_24 (Dense)            (None, 4)                 262148

=================================================================
Total params: 1,434,404
Trainable params: 1,434,404
Non-trainable params: 0
_________________________________________________________________
None
```

Figure 4-1-2 : CNN model for skin classification using categorical cross-entropy

The batch size and the No. of epochs are very important to increase the accuracy of the model. Batch size represents the images being processed at a time. The batch size chosen is 400. The model is trained multiple times to achieve this accuracy.

Multiple things are focused to increase the accuracy which are the Best Model, Best Learning rate, Best Optimizer, Large training data, Batch size, Number of Epochs and

Multiple trainings with multiple parametric settings. The results of the accuracy of the model is shown below.

The accuracy of the model is 0.85 x 100% = 85 % using Epoch = 50.

```
epoch=50  # change epoch number for good results
hist=model.fit(X_train, y_train, batch_size=400, epochs=epoch, validation_data=(X_test, y_test))
```

```
Epoch 40/50
4/4 [==============================] - 5s 1s/step - loss: 0.0023 - accuracy: 0.9993 - val_loss: 1.6490 - val_accuracy: 0.8576
Epoch 41/50
4/4 [==============================] - 5s 1s/step - loss: 0.0020 - accuracy: 0.9993 - val_loss: 1.6395 - val_accuracy: 0.8543
Epoch 42/50
4/4 [==============================] - 5s 1s/step - loss: 0.0013 - accuracy: 1.0000 - val_loss: 1.6492 - val_accuracy: 0.8576
Epoch 43/50
4/4 [==============================] - 5s 1s/step - loss: 4.7657e-04 - accuracy: 1.0000 - val_loss: 1.6837 - val_accuracy: 0.8543
Epoch 44/50
4/4 [==============================] - 5s 1s/step - loss: 7.6166e-04 - accuracy: 1.0000 - val_loss: 1.6797 - val_accuracy: 0.8510
Epoch 45/50
4/4 [==============================] - 5s 1s/step - loss: 3.2729e-04 - accuracy: 1.0000 - val_loss: 1.6673 - val_accuracy: 0.8543
Epoch 46/50
4/4 [==============================] - 5s 1s/step - loss: 2.5544e-04 - accuracy: 1.0000 - val_loss: 1.6593 - val_accuracy: 0.8477
Epoch 47/50
4/4 [==============================] - 5s 1s/step - loss: 2.1452e-04 - accuracy: 1.0000 - val_loss: 1.6554 - val_accuracy: 0.8477
Epoch 48/50
4/4 [==============================] - 5s 1s/step - loss: 1.8446e-04 - accuracy: 1.0000 - val_loss: 1.6551 - val_accuracy: 0.8510
Epoch 49/50
4/4 [==============================] - 5s 1s/step - loss: 1.5328e-04 - accuracy: 1.0000 - val_loss: 1.6589 - val_accuracy: 0.8510
Epoch 50/50
4/4 [==============================] - 5s 1s/step - loss: 1.3110e-04 - accuracy: 1.0000 - val_loss: 1.6655 - val_accuracy: 0.8510
```

Figure 4-1-3 : CNN Model accuracy results

The first graph shows the Train Loss versus Validation Loss while the second graph shows the Train Accuracy versus Validation Accuracy.

```
[0.501655638217926, 0.45860928297042847, 0.5496688485145569, 0.55
```

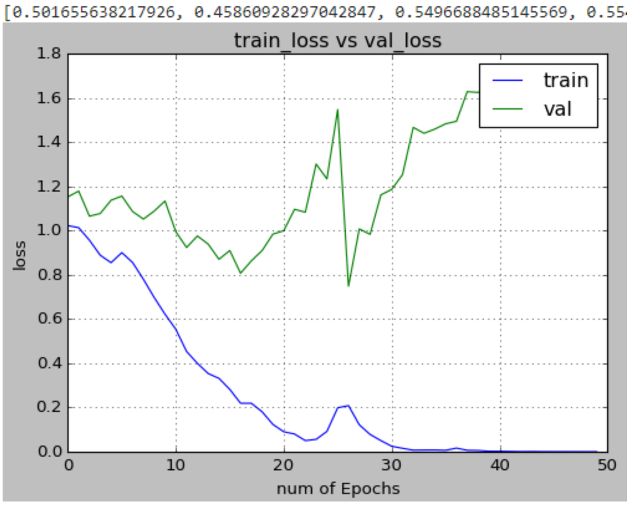


Figure 4-1-4 : Graph of Train Loss vs Validation Loss

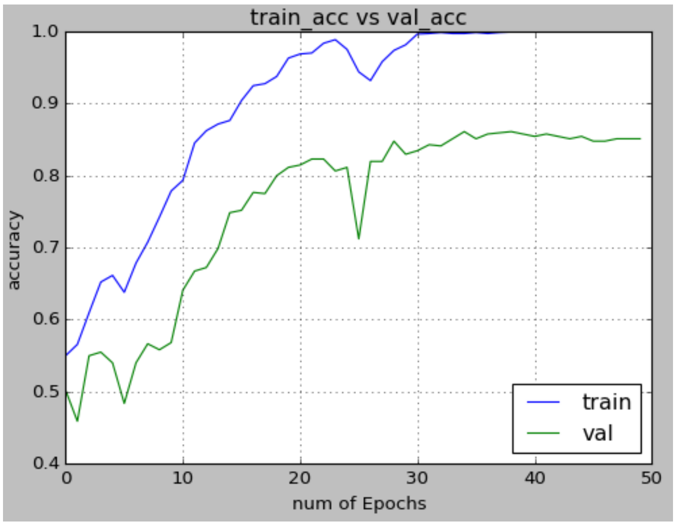


Figure 4-1-5 :Graph of Train Accuracy vs Validation Accuracy

The model has reached the validation loss of 1.6 and training loss of 0.0001314. The training accuracy of the model is 1.0000 while the validation accuracy is 0.8510.

An image is taken from the dataset under combination skin type category for prediction. The results are accurate and show that it is categorized as combination skin type.

```
[[0.28766206 0.25490537 0.19023265 0.2671999 ]]
[0]
Text(0.5, 1.0, 'Skin Type\n Combination')
```

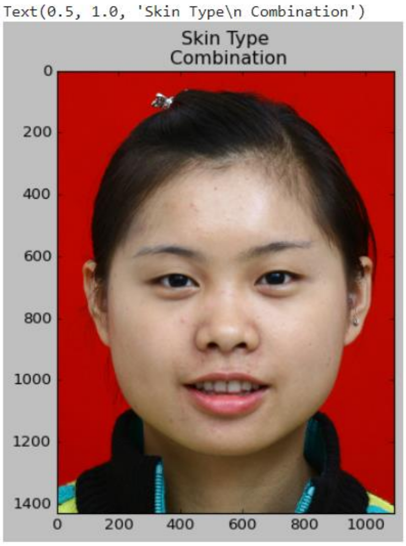


Figure 4-1-6 : Combination Skin Type image

4-2 Sentiment analysis for DistilBERT model

The dataset is imported into the model.

	product_id	review_title	review_text	rating	age_range	skin_type	skin_tone	eye_color	reviewer_username	tags	review_id
0	P38217	Worth the money	Sometimes I stray from this cleanser, but I al...	5.0	NaN	normal	light	NaN	katechatte	{foamy,exfoliating}	6611717f-2636-4756-bf36-66c81cc267a7
1	P38217	Great	I am a 41 year old African American woman with...	5.0	NaN	combination	deep	NaN	snook41	{foamy}	e7d3307e-02ff-45a1-8fc3-6bd628bedd86
2	P38217	Great Product	I'm really enjoying this product. Received a s...	5.0	NaN	combination	olive	NaN	wahinewarrior	{foamy,milky,exfoliating}	4188d728-fde6-4d06-984e-164cca2b8781
3	P38217	Nice, but not great for combination skin	I tried this cleanser at a friends house, and ...	3.0	NaN	combination	fair	NaN	jenlines22	{hydrating,creamy}	248c904c-6e30-4929-8228-87b03ad7a921
4	P38217	great moisturizer	leaves the skin feeling fresh and revived... j...	5.0	NaN	dry	light	NaN	jessea	{exfoliating}	654bdb99-9371-4440-a540-0dd2a73da339

Figure 4-2-1 : Imported Dataset

The sentiment labels is generated from the data using the ratings given by each user. The reviews that have a rating of less than 3 are given negative sentiment labels and those having 3 or more than 3 are given positive sentiment labels. For training purposes, 'positive' and 'negative' labels are mapped to 1 and 0 respectively.

	review_text	sentiment
0	Sometimes I stray from this cleanser, but I al...	1
1	I am a 41 year old African American woman with...	1
2	I'm really enjoying this product. Received a s...	1
3	I tried this cleanser at a friends house, and ...	1
4	leaves the skin feeling fresh and revived... j...	1

Figure 4-2-2: Sentiment labels

Observing the distribution of labels in the data, it is shown that there are far more positive reviews than negative reviews. During training, this might induce bias into the model. To counter this, a text augmentation method is applied where new texts are generated from a given text which has the same semantics and meaning but will just have a different framing. This text augmentation is applied to the negative reviewed texts only since there are very less number of negative reviewed texts as compared to positive reviewed texts. The back translation technique is used which initially converts

English text to German and then back to English.

```
Downloading: 100% 825/825 [00:00<00:00, 18.7kB/s]
Downloading: 100% 1.08G/1.08G [00:33<00:00, 33.2MB/s]
Downloading: 100% 825/825 [00:00<00:00, 19.1kB/s]
Downloading: 100% 1.08G/1.08G [00:33<00:00, 32.2MB/s]
Downloading: 100% 67.0/67.0 [00:00<00:00, 1.53kB/s]
Downloading: 100% 849k/849k [00:01<00:00, 826kB/s]
Downloading: 100% 849k/849k [00:00<00:00, 691kB/s]
Downloading: 100% 315k/315k [00:00<00:00, 364kB/s]
Downloading: 100% 67.0/67.0 [00:00<00:00, 1.79kB/s]
Downloading: 100% 849k/849k [00:00<00:00, 1.03MB/s]
Downloading: 100% 849k/849k [00:00<00:00, 999kB/s]
Downloading: 100% 315k/315k [00:00<00:00, 280kB/s]
```

Figure 4-2-3: Text Augmentation using Back Translation Technique

It is shown that the augmented text basically means the same as the original text. Only the framing of the sentence is changed a bit in the augmented text. This technique to all the negatively reviewed sentences.

```
Original Text:  I'm really enjoying this product. Received a sample trio of the Murad products to try first, then purchased the cleanser. Great deal at $35 as a little goes a long way! I use it nightly with

/usr/local/lib/python3.7/dist-packages/torch/_tensor.py:575: UserWarning: floor_divide is deprecated, and will be removed in a future version of pytorch. It currently rounds toward 0 (like the 'trunc' funct
To keep the current behavior, use torch.div(a, b, rounding_mode='trunc'), or for actual floor division, use torch.div(a, b, rounding_mode='floor'). (Triggered internally at  /pytorch/aten/src/ATen/native/Bi
  return torch.floor_divide(self, other)
Augmented Text:  I got a sample trio of Murad products to try first, then I bought the cleanser. A lot for $35, because a bit much works! I use it every night with my Clarisonic Mia and in the morning alone
```

Figure 4-2-4: Example of Text Augmentation

After performing text augmentation on the negatively reviewed texts, it is shown that number of texts with negative (with 0 labels) has doubled. This will certainly help in training the sentiment analysis model better and improve the accuracy of the model. Next, the data processed and generated is used to train a sentiment analysis model. The DistilBERT pre-trained model is fine-tuned according to the use case using the huggingface transformers library.

```
Downloading: 100% 232k/232k [00:00<00:00, 1.13MB/s]
Downloading: 100% 48.0/48.0 [00:00<00:00, 1.33kB/s]
Downloading: 100% 629/629 [00:00<00:00, 21.2kB/s]
Text:  I have been using this product for almost a month in the morning. ( I am also using the Kate Somerville nourish in the evenings after cleanser). This product line was recommended to me by the skin care consultant, after
Tokenized Text:  {'input_ids': [[101, 1045, 2031, 2042, 2478, 2023, 4031, 2005, 2471, 1037, 3204, 1999, 1996, 2851, 1012, 1006, 1045, 2572, 2036, 2478, 1996, 5736, 27209, 2053, 9496, 4095, 1999, 1996, 16241, 2044, 4550, 8043,
```

Figure 4-2-5: Tokenized Text

```
Downloading: 100%                                          268M/268M [00:06<00:00, 37.4MB/s]
All model checkpoint layers were used when initializing TFDistilBertForSequenceClassification.

All the layers of TFDistilBertForSequenceClassification were initialized from the model checkpoint at distilbert-base-uncased-finetuned-sst-2-english.
If your task is similar to the task the model of the checkpoint was trained on, you can already use TFDistilBertForSequenceClassification for predictions without further training.
Model: "tf_distil_bert_for_sequence_classification"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
distilbert (TFDistilBertMain multiple                  66362880
_________________________________________________________________
pre_classifier (Dense)       multiple                  590592
_________________________________________________________________
classifier (Dense)           multiple                  1538
_________________________________________________________________
dropout_19 (Dropout)         multiple                  0
=================================================================
Total params: 66,955,010
Trainable params: 66,955,010
Non-trainable params: 0
_________________________________________________________________
```

Figure 4-2-6: DistilBERT model for sentiment analysis

The results of the accuracy of the model is shown below. The accuracy of the model is 0.95 x 100% = 95% using Epoch = 10. The dataset can be increased and the value of epoch can be increased to increase the accuracy of the model.

```
Epoch 1/10
WARNING:tensorflow:The parameters `output_attentions`, `output_hidden_states` and `use_cache` cannot be updated when calling a model.T
WARNING:tensorflow:AutoGraph could not transform <bound method Socket.send of <zmq.Socket(zmq.PUSH) at 0x7f1ff3f30ec0>> and will run i
Please report this to the TensorFlow team. When filing the bug, set the verbosity to 10 (on Linux, `export AUTOGRAPH_VERBOSITY=10`) an
Cause: module, class, method, function, traceback, frame, or code object was expected, got cython_function_or_method
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING: AutoGraph could not transform <bound method Socket.send of <zmq.Socket(zmq.PUSH) at 0x7f1ff3f30ec0>> and will run it as-is.
Please report this to the TensorFlow team. When filing the bug, set the verbosity to 10 (on Linux, `export AUTOGRAPH_VERBOSITY=10`) an
Cause: module, class, method, function, traceback, frame, or code object was expected, got cython_function_or_method
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING:tensorflow:AutoGraph could not transform <function wrap at 0x7f200f0e5950> and will run it as-is.
Cause: while/else statement not yet supported
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING: AutoGraph could not transform <function wrap at 0x7f200f0e5950> and will run it as-is.
Cause: while/else statement not yet supported
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convertWARNING:tensorflow:The parameter `return_

WARNING:tensorflow:The parameters `output_attentions`, `output_hidden_states` and `use_cache` cannot be updated when calling a model.T
WARNING:tensorflow:The parameter `return_dict` cannot be set in graph mode and will always be set to `True`.
50/50 [==============================] - ETA: 0s - loss: 0.3495 - accuracy: 0.8938WARNING:tensorflow:The parameters `output_attentions
WARNING:tensorflow:The parameter `return_dict` cannot be set in graph mode and will always be set to `True`.
50/50 [==============================] - 569s 10s/step - loss: 0.3495 - accuracy: 0.8938 - val_loss: 0.1739 - val_accuracy: 0.9400
Epoch 2/10
50/50 [==============================] - 520s 10s/step - loss: 0.1678 - accuracy: 0.9375 - val_loss: 0.1520 - val_accuracy: 0.9500
Epoch 3/10
50/50 [==============================] - 520s 10s/step - loss: 0.1145 - accuracy: 0.9575 - val_loss: 0.1514 - val_accuracy: 0.9500
Epoch 4/10
50/50 [==============================] - 528s 11s/step - loss: 0.0695 - accuracy: 0.9775 - val_loss: 0.1612 - val_accuracy: 0.9550
Epoch 5/10
50/50 [==============================] - 520s 10s/step - loss: 0.0510 - accuracy: 0.9850 - val_loss: 0.1694 - val_accuracy: 0.9550
Epoch 6/10
50/50 [==============================] - 521s 10s/step - loss: 0.0266 - accuracy: 0.9925 - val_loss: 0.1859 - val_accuracy: 0.9500
Epoch 7/10
50/50 [==============================] - 525s 11s/step - loss: 0.0085 - accuracy: 1.0000 - val_loss: 0.1951 - val_accuracy: 0.9500
Epoch 8/10
50/50 [==============================] - 513s 10s/step - loss: 0.0102 - accuracy: 0.9975 - val_loss: 0.2111 - val_accuracy: 0.9450
Epoch 9/10
50/50 [==============================] - 515s 10s/step - loss: 0.0071 - accuracy: 0.9987 - val_loss: 0.2349 - val_accuracy: 0.9400
Epoch 10/10
50/50 [==============================] - 514s 10s/step - loss: 0.0024 - accuracy: 1.0000 - val_loss: 0.2274 - val_accuracy: 0.9500
<keras.callbacks.History at 0x7f1f30886290>

3h 53m 40s   completed at 4:57 PM
```

Figure 4-2-7: Binary Cross-entropy for sentiment analysis

```
WARNING:tensorflow:The parameters `output_attentions`, `output_hidden_states` and `use_cache` cannot be updated when calling a model.They have to be set to True/False in the config object (i.e.: `config=XConfig.from_pretrained('name', output_attent
WARNING:tensorflow:AutoGraph could not transform <bound method Socket.send of <zmq.Socket(zmq.PUSH) at 0x7f5106786ec0>> and will run it as-is.
Please report this to the TensorFlow team. When filing the bug, set the verbosity to 10 (on Linux, `export AUTOGRAPH_VERBOSITY=10`) and attach the full output.
Cause: module, class, method, function, traceback, frame, or code object was expected, got cython_function_or_method
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING: AutoGraph could not transform <bound method Socket.send of <zmq.Socket(zmq.PUSH) at 0x7f5106786ec0>> and will run it as-is.
Please report this to the TensorFlow team. When filing the bug, set the verbosity to 10 (on Linux, `export AUTOGRAPH_VERBOSITY=10`) and attach the full output.
Cause: module, class, method, function, traceback, frame, or code object was expected, got cython_function_or_method
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING:tensorflow:AutoGraph could not transform <function wrap at 0x7f512193b950> and will run it as-is.
Cause: while/else statement not yet supported
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING: AutoGraph could not transform <function wrap at 0x7f512193b950> and will run it as-is.
Cause: while/else statement not yet supported
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
WARNING:tensorflow:The parameter `return_dict` cannot be set in graph mode and will always be set to `True`.
WARNING:tensorflow:The parameters `output_attentions`, `output_hidden_states` and `use_cache` cannot be updated when calling a model.They have to be set to True/False in the config object (i.e.: `config=XConfig.from_pretrained('name', output_attent
WARNING:tensorflow:The parameter `return_dict` cannot be set in graph mode and will always be set to `True`.
WARNING:tensorflow:The parameters `output_attentions`, `output_hidden_states` and `use_cache` cannot be updated when calling a model.They have to be set to True/False in the config object (i.e.: `config=XConfig.from_pretrained('name', output_attent
WARNING:tensorflow:The parameter `return_dict` cannot be set in graph mode and will always be set to `True`.


For Product 2, out of 38 people with dry skin type, 89.47%  posted positive reviews while 10.53% posted negative reviews.

For Product 2, out of 5 people with oily skin type, 100.0%  posted positive reviews while 0.0% posted negative reviews.

For Product 2, out of 7 people with normal skin type, 100.0%  posted positive reviews while 0.0% posted negative reviews.

For Product 2, out of 40 people with combination skin type, 92.5%  posted positive reviews while 7.5% posted negative reviews.
```

Figure 4-2-8: Sentiment analysis of product 2

After training the model, it is saved and loaded to analyse the reviews. For example when product 2 is input in the model, the percentage of positive and negative reviews for each of the skin type is shown. Therefore, the next step can be proceeded where the product with high percentage positive reviews is recommended.

A sample review text is retrieved and predicted. The sentiment labelled negative results is shown and proves the accuracy of model.

```
WARNING:tensorflow:The parameters `output_attentions`, `output_hidden_states` and `use_cach
WARNING:tensorflow:The parameter `return_dict` cannot be set in graph mode and will always
TEXT:  I have dry skin and I got a sample of this product, I tried it, but it didn't work.

SENTIMENT:  Negative
```

Figure 4-2-9: Sample Text Prediction

4-3 Skin Care Product Recommendation System

```
get_product_recommendation(skin_type="normal")
```

```
'Product 1'
```

Figure 4-3-1 : Product recommendation for skin type 'natural'

A threshold is set when the total amount of reviews of product for the skin type has to be more then 15 and has more than 86% positive reviews, only it will be recommended. For example, when the skin type 'dry' is input, 'product 3' is recommended.

The CNN model for image classification and DistilBERT model is combined and becomes a recommendation system. When a random image is input into the recommendation system, it will firstly determines the skin type. Then the product is recommended according to the skin type.

```
recommend_product_from_img('/content/mydrive/MyDrive/Train/dry skin.jpg') # input image
```

```
'Recommended Product for detected skin type dry: Product 3'
```

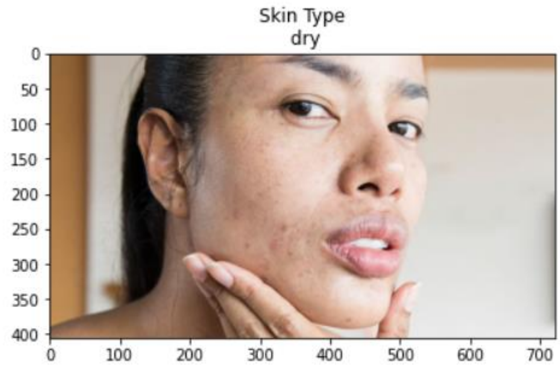


Figure 4-3-2: Experimental results for dry skin type

For example, when the image with dry skin type is input in the recommendation system, the results 'dry' is shown and 'Product 3' is recommended.

4-4 Telegram Chatbot for User Interaction

In the proposed project, the Telegram Chatbot connects the user and the skin care product recommendation system. User can easily use the chatbot to get recommendations for the beauty skin care products. The Telegram Chatbot will ask the name of the user and greet the user after the name is provided. Then the chatbot will request the user to input his/her facial image to the chatbot for skin identification and product recommendation. There are no limitations of the amount of facial images the users input, they can input as many facial images as they want.

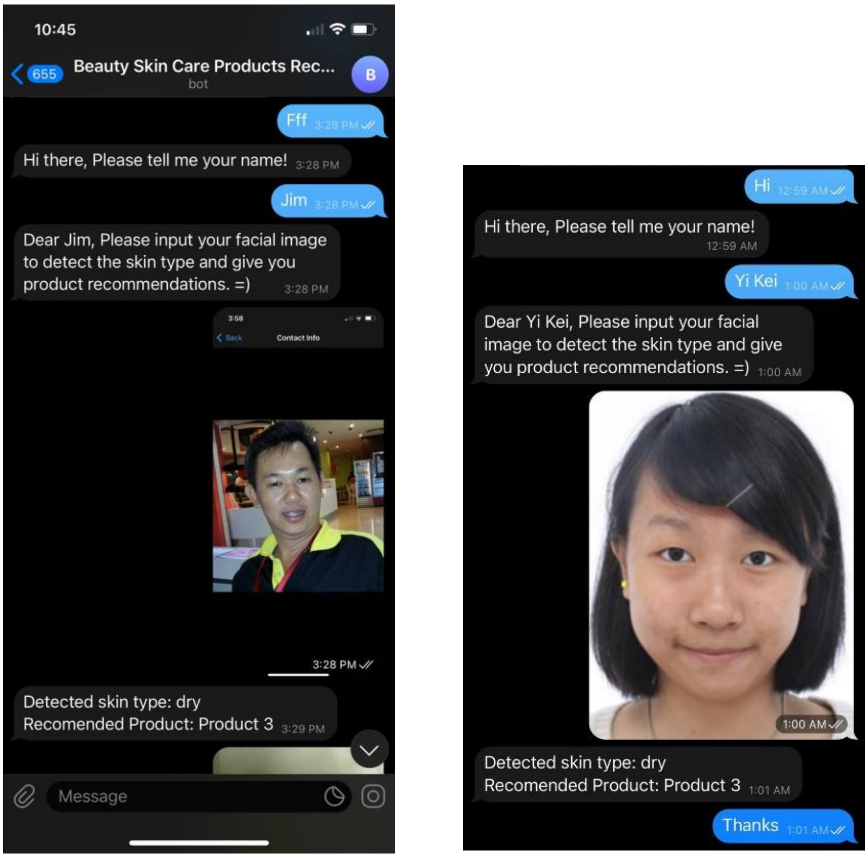


Figure 4-4-1: Screenshot of Chatbot in proposed system

If the image of the users input is not a facial image, the system is unable to detect the face of the user and chatbot will give the response of "No face detected".

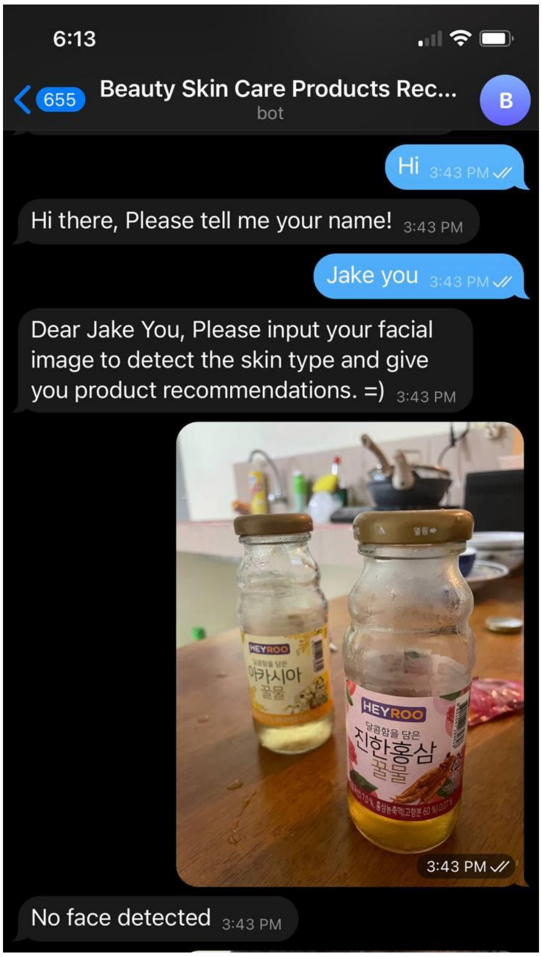


Figure 4-4-2: Screenshot of Chatbot in proposed system

Users can access Beauty Skin Care Product Chatbot can through many ways with the existing Telegram app:

1) via Telegram user search function with the username @BeautySkinCare_Chatbot
2) via the link below: https://t.me/BeautySkinCare_Chatbot
3) Scan the QR code below:

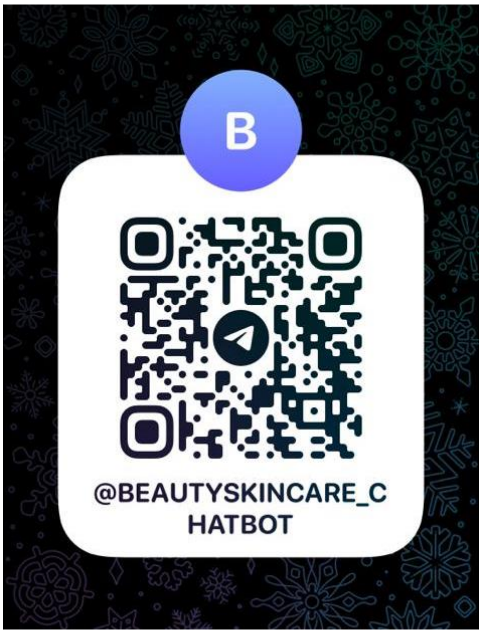


Figure 4-4-3: Screenshot of QR code of Chatbot

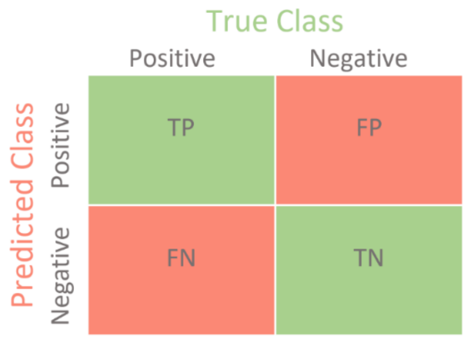
CHAPTER 5 – SYSTEM EVALUATION AND DISCUSSION

5-1 System Testing and Performance Metrics

System Testing is performed for the chatbot and the performance metrics is used to evaluate the performance of the chatbot. The facial images with the known skin type identification is sent to the chatbot.

5-1-1 Confusion Matrix

Confusion Matrix is the summary of model prediction results on a classification problem. It can be used for the performance metrics of chatbot where the number of correct and incorrect skin identification predictions are summarized by different classes. It is a N X N matrix where N is the number of skin type classes being predicted. We have four classes which are oily, dry, combination and natural. Therefore N = 4 and we have 4 X 4 matrix.



TP(True positive): Observation is positive, and is predicted positive.

TN(True negative): Observation is negative, and is predicted negative.

FP(False positive): Observation is negative, and is predicted positive.

FN(False negative): Observation is positive, and is predicted negative.

Figure 5-1-1: Confusion Matrix

The four classes of skin type classification are oily, dry, combination and natural with 100 datasets for every class. The TP value is counted when the actual value and predicted value is the same. It means when the actual skin type image matches with the prediction of the chatbot skin identification result, the TP value will be counted. For example, if the combination skin image is input in the telegram chatbot, and the skin identification results of the chatbot is also combination skin type (refer *Figure 5-1-3)*, then it counts as TP (True positive). The TN, FP, and FN values are counted respectively with the addition of cells, but with the Confusion Matrix Online Calculator, we do not have to count them manually. (refer *Figure 5-1-4)*

Figure 5-1-2: Combination skin image

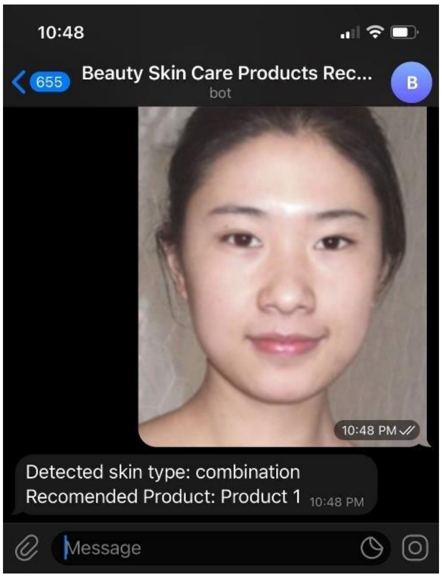


Figure 5-1-3: Screenshot of Chatbot

The results show that the overall accuracy of the evaluation metrics of chatbot is 81.25%. The precision is also called the positive prediction value which is the proportion of positive cases that are identified correctly. The precision of the model evaluation for the classes are 0.85, 0.78, 0.78, 0.83 respectively. The recall is also called sensitivity which is the proportion of actual positive cases correctly identified. The recall of the model evaluation for the classes are 0.88, 0.83, 0.77,0.78 respectively. The F1 Score is the harmonic mean of precision and recall values for the classification problem. F1 Score helps measure Recall and Precision at the same time. [16] The F1 score of the model evaluation for the classes are 0.86, 0.80, 0.78, 0.80.

Class 1 : Dry Skin Type

Class 2 : Oily Skin Type

Class 3 : Combination Skin Type

Class 4 : Natural Skin Type

Number of classes: - 4 +

	Class 1	Class 2	Class 3	Class 4
Class 1	85	5	3	7
Class 2	7	78	11	4
Class 3	2	6	79	13
Class 4	3	5	9	83
Total for Class	97	94	102	107

No cell selected

Results

TP: 325

Overall Accuracy: 81.25%

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
1	97	100	93.25%	0.85	0.88	0.86
2	94	100	90.5%	0.78	0.83	0.80
3	102	100	89%	0.79	0.77	0.78
4	107	100	89.75%	0.83	0.78	0.80

Figure 5-1-4: Results of Confusion Matrix Online Calculator

5-2 Project Challenges

The biggest challenge for the project is finding the suitable dataset, and the best accuracy affecting parameters (Learning rate, batch size, number of epochs). The number of data is increased to 1000 with the number of 250 data per skin type data set. The dataset is manually collected from various websites one by one, therefore it is very time consuming. The data is then analyzed and labelled manually, which needs a very big effort to do so. Then data was increased making a copy of 250 images per skin type with little variations. The actual dataset is 1000 but the overall images used to train the model is 2000. The model is retrained again and again to increase the accuracy. It has high difficulty to achieve high accuracy. The accuracy of CNN model reaches 85% after training the model multiple times with multiple parameters.

5-3 Objective Evaluation

The first objective is to train a skin type classification model using CNN for skin type based product recommendations. The accuracy of CNN model is 85% using Epoch = 50.The second objective is to train a sentiment analysis model using product reviews data crawled from cosmetic website. The accuracy of the model is 95% using Epoch = 10.The last objective is to train a recommendation system to suggest product types based on user skin types. The overall accuracy of the evaluation metrics of chatbot is 81.25%.

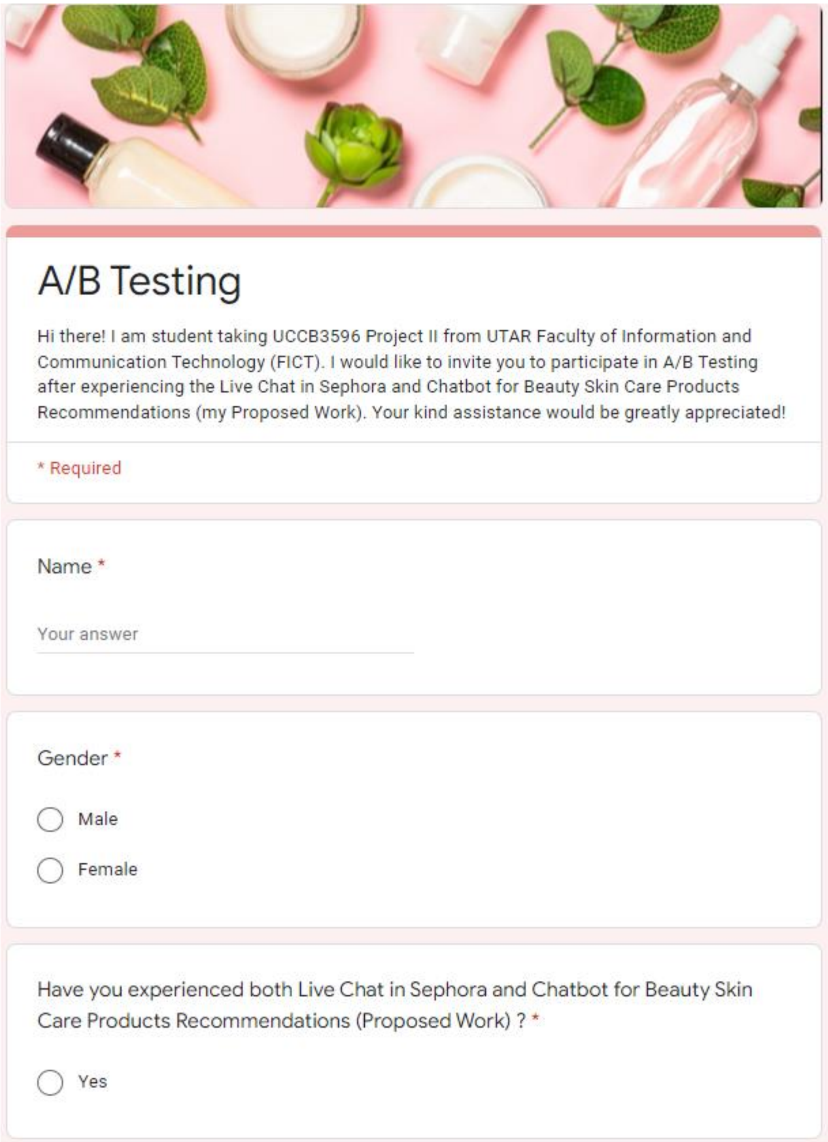
CHAPTER 6 – A/B TESTING

6-1 Google Forms for A/B Testing

The google forms is distributed to the participants after experiencing both Live Chat in Sephora and Chatbot for Beauty Skin Care Products Recommendations (Proposed Work).

Link of google forms :

https://docs.google.com/forms/d/e/1FAIpQLSerKNwV9JniDtH6dhXZ8SbU3kHHHwyD0VmzOPBSn05b0XSi6g/viewform?usp=pp_url



A/B Testing

Hi there! I am student taking UCCB3596 Project II from UTAR Faculty of Information and Communication Technology (FICT). I would like to invite you to participate in A/B Testing after experiencing the Live Chat in Sephora and Chatbot for Beauty Skin Care Products Recommendations (my Proposed Work). Your kind assistance would be greatly appreciated!

* Required

Name *

Your answer

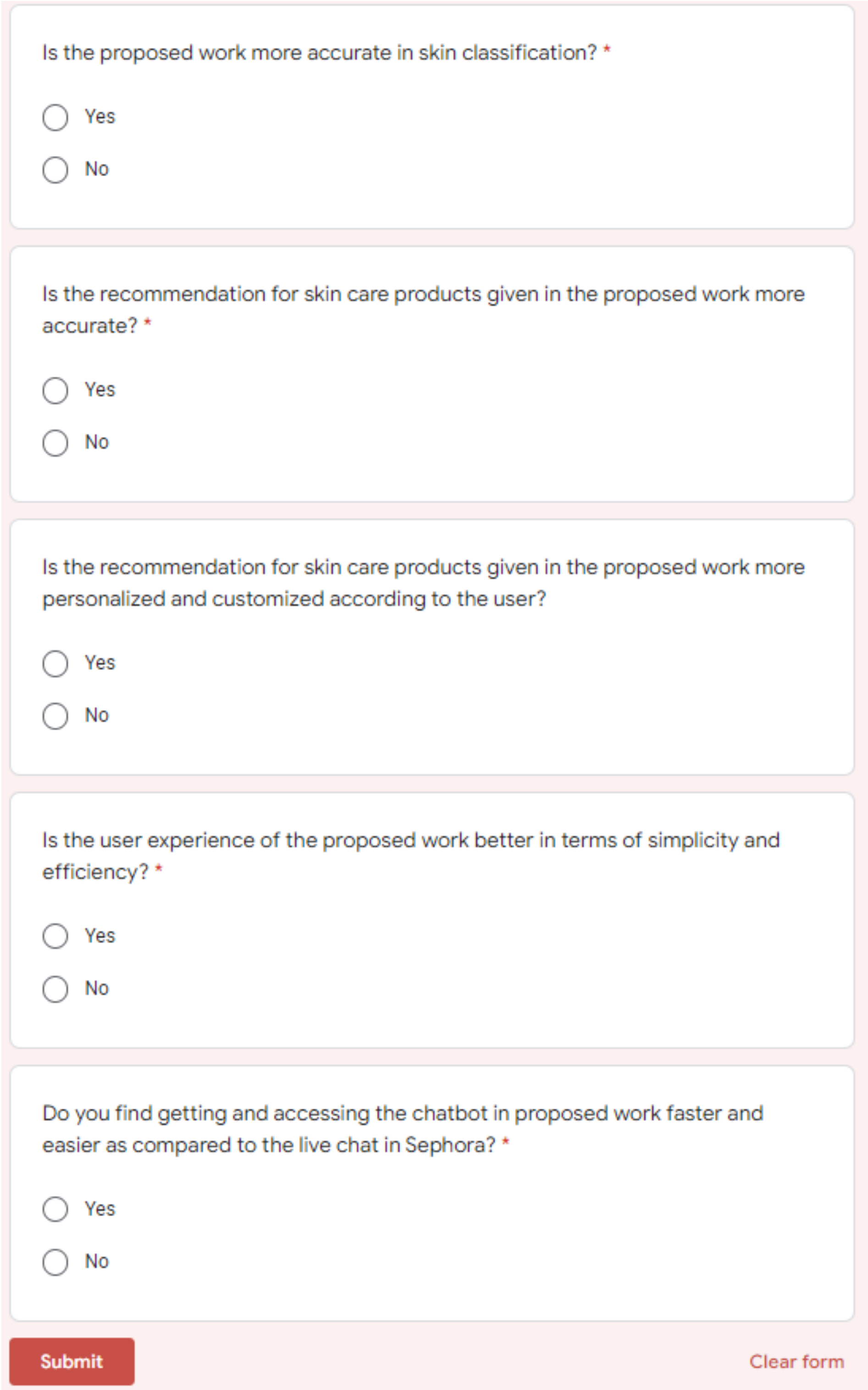
Gender *

Male

Female

Have you experienced both Live Chat in Sephora and Chatbot for Beauty Skin Care Products Recommendations (Proposed Work) ? *

Yes



Is the proposed work more accurate in skin classification? *

Yes

No

Is the recommendation for skin care products given in the proposed work more accurate? *

Yes

No

Is the recommendation for skin care products given in the proposed work more personalized and customized according to the user?

Yes

No

Is the user experience of the proposed work better in terms of simplicity and efficiency? *

Yes

No

Do you find getting and accessing the chatbot in proposed work faster and easier as compared to the live chat in Sephora? *

Yes

No

Submit

Clear form

Figure 6-1: Screenshot of Google Forms for A/B Testing

6–2 Analyse Results for A/B Testing

1. 20 respondents with equally 10 males and 10 females participated in the A/B Testing.

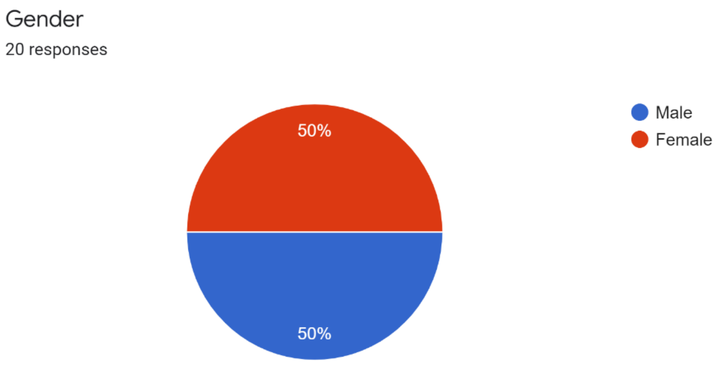


Figure 6-2-1: Pie chart Results for A/B Testing

2. All the respondents have experienced both Live Chat in Sephora and Chatbot for Beauty Skin Care Products Recommendations (Proposed Work)

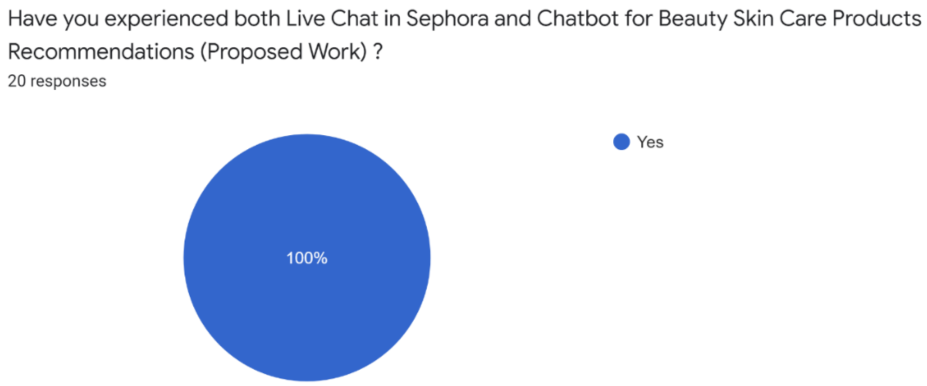


Figure 6-2-2: Pie chart Results for A/B Testing

3. The results show that there are more respondents (90%) who feel the proposed work is more accurate in skin classification. This is because the proposed work can identify the skin type of the user when the facial image of the user is sent to the chatbot. As compared to the Sephora Live Chat (refer to *Figure 6-2-4*) who does not have the feature online to determine skin type, the proposed chatbot can detect the skin type easily and recommend products to user based on their skin type. (refer to *Figure 6-2-5*).

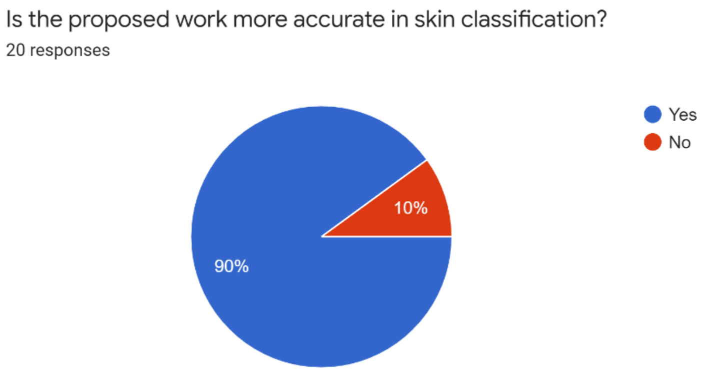


Figure 6-2-3: Pie chart Results for A/B Testing

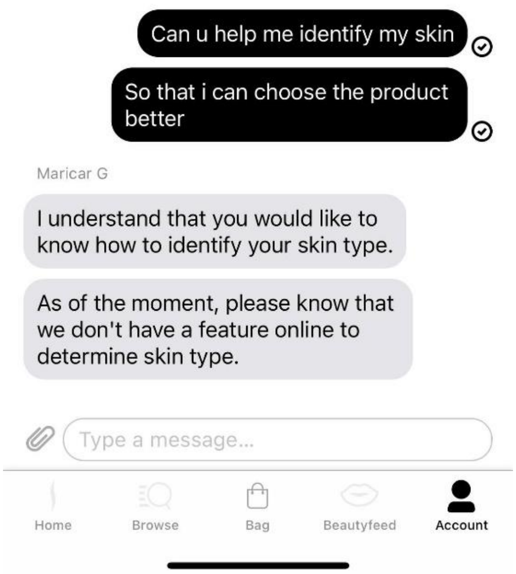


Figure 6-2-4: Screenshot of Live Chat in Sephora

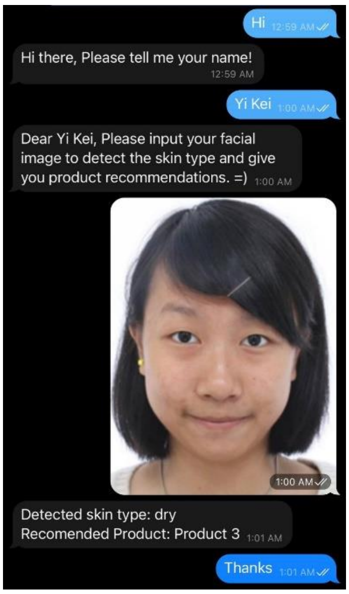


Figure 6-2-5: Screenshot of Chatbot in proposed work

4. 85% of the respondents feel that the recommendation for skin care products given in proposed work is more accurate compared to the live chat in Sephora with only 15% of the respondents. The proposed work helps identify the skin type of the users and give specific product recommendations while Sephora recommends the best-selling skincare products (refer *Figure 6-2-7*). Sephora recommends the skin care products based on the popularity of the products but not according to the skin type of the users. Therefore, the respondents feel that the skin care product recommendations for Sephora is not as accurate as compared to the proposed work.

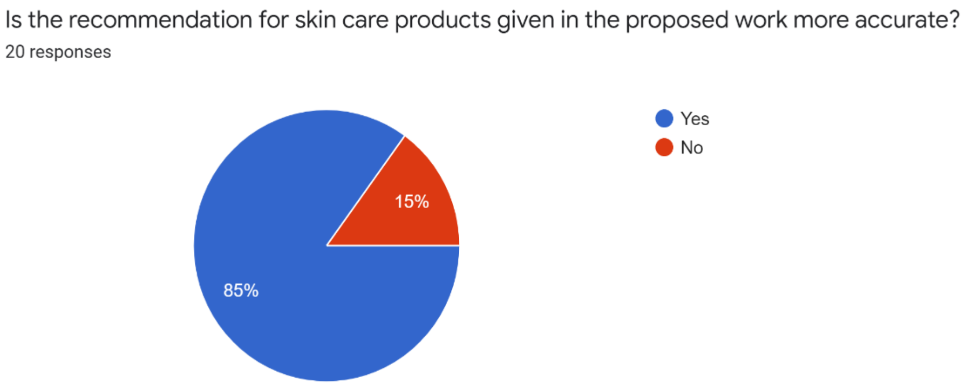


Figure 6-2-6: Pie chart Results for A/B Testing

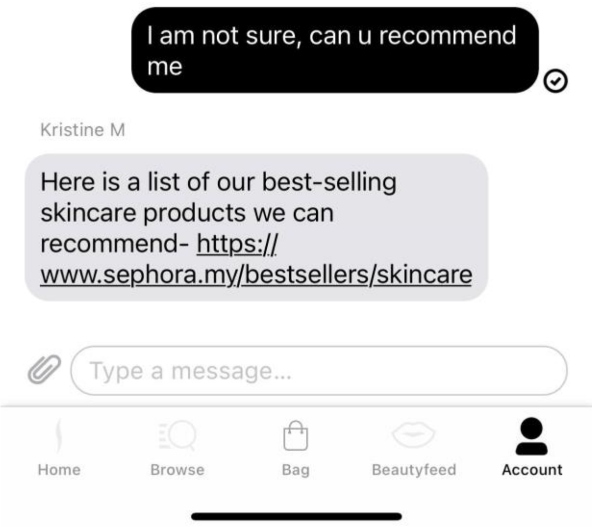


Figure 6-2-7: Screenshot of Live Chat in Sephora

5. There are more respondents with 90% of them who feel that the proposed chatbot gives more personalized and customized skin care products recommendations according to the user while Sephora provides various skin care products for users to choose himself/herself (refer *Figure 6-2-9*). In Sephora, users have to browse the link provided and go through all of the skin care products before purchasing the product. Meanwhile the proposed chatbot will give personalized recommendations to the users based on his/her skin type.

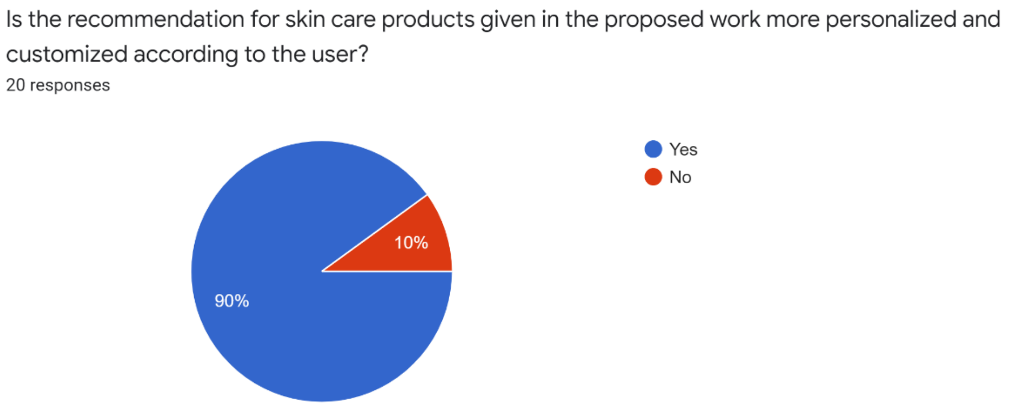


Figure 6-2-8: Pie chart Results for A/B Testing

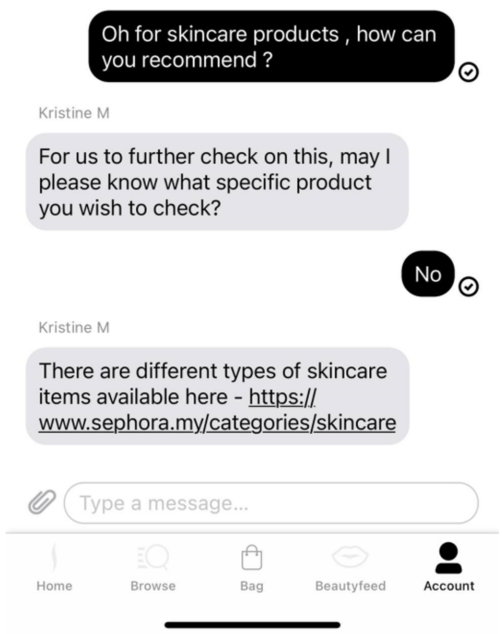


Figure 6-2-9: Screenshot of Live Chat in Sephora

6. All the respondents have better user experience using the proposed work in terms of simplicity and efficiency. The proposed work is simple to be used and it is straight forward as it provides the product recommendation after the user inputs his/her facial image. Whereas the live chat in Sephora asks a bunch of questions which is time consuming and does not serve a big purpose (refer *Figure 6-2-11*). The live chat in Sephora recommends the skin care products which are top rated solely, which is not accurate and efficient. It also assumes that the user has combination skin type and gives product recommendations when the user cannot identify the skin type himself/herself (refer *Figure 6-2-12*).

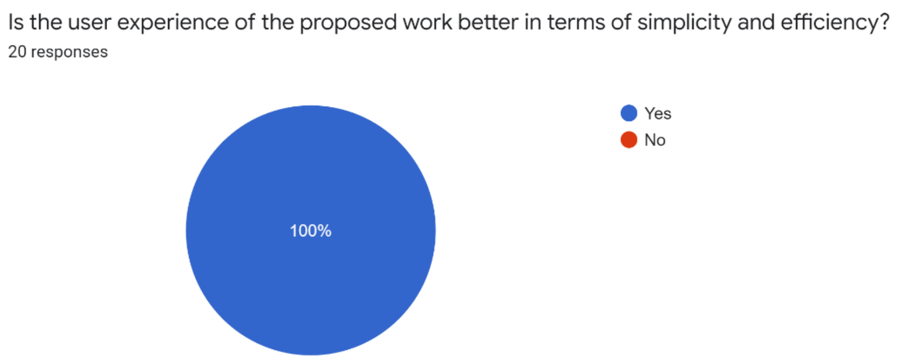


Figure 6-2-10: Pie chart Results for A/B Testing

Figure *6-2-11*: Screenshot of Live Chat in Sephora

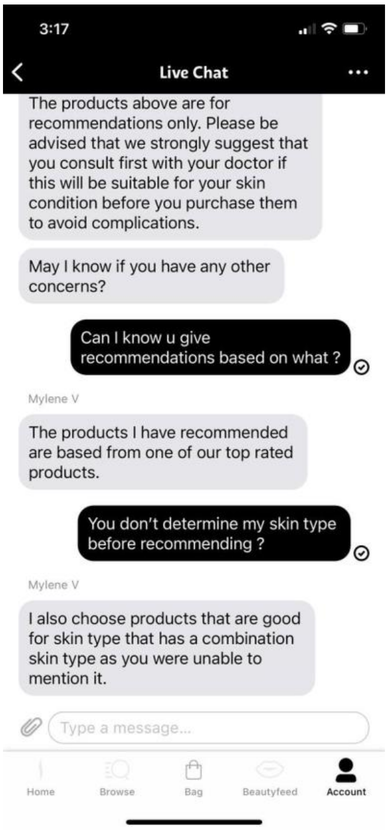


Figure *6-2-12*: Screenshot of Live Chat in Sephora

7. There are more respondents with 95% of them who find accessing the proposed work faster and easier. Unlike the Sephora application, users have to download the app through Google play store or iOS app store. (refer *Figure 6-2-14)* In the proposed work, users just need to search the username of the chatbot via Telegram and does not need to download any apps. (refer *Figure 6-2-15)* The proposed work provides easier accessibility and convenience for the users so that they can access the chatbot in a short time.

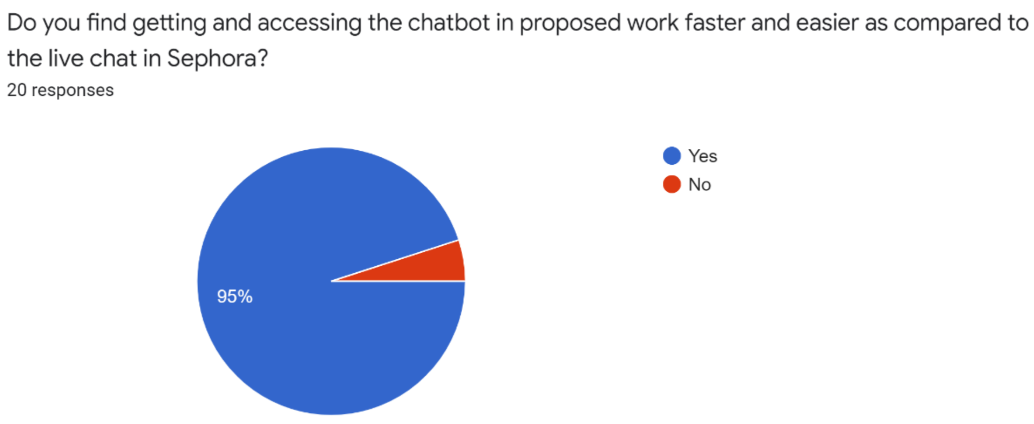


Figure *6-2-13*: Pie chart Results for A/B Testing

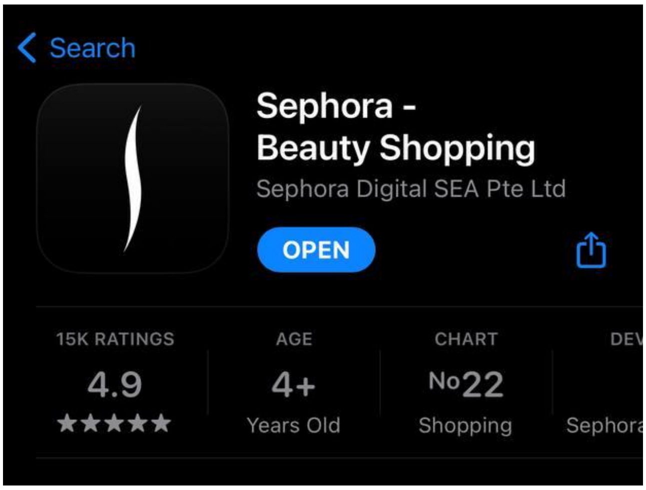


Figure 6-2-14: Screenshot of Sephora App

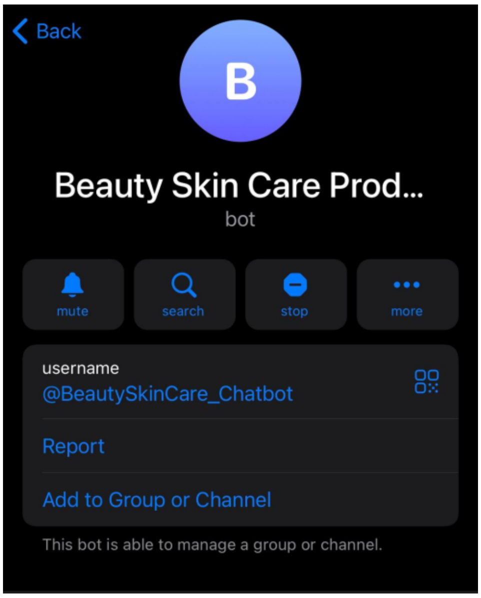


Figure 6-2-15: Screenshot of Proposed Chatbot

CHAPTER 7 - CONCLUSION

Various chatbots are developed to recommend beauty skincare products to the customer. This project's problem domain is the chatbots that exist now suggest the products to users based on the product description. The existing chatbots make the users' life easier by recommending products to users based on the product description labelled on the products. However, the product description might not be accurate because it is labelled by the manufacturer. Furthermore, the product might not suite the users with different skin types. In this project, a real-world review is used for chatbots to recommend products to other related users according to their skin type.

The novelty of the project is that the proposed machine learning model is based on user reviews and predictions are made based on real reviews using sentiment analysis according to the user's particular skin type. Due to the fact that different product is suitable for different skin type users, the user review plays an important role for sentiment analysis. The reviews are from users that have experienced the products before and give positive or negative reviews to the product. The machine learning model uses training data that are preprocessed and accurate. By inputting the image of portrait, the skin type of user can be determined and the beauty skin care products are recommended to the user. The speciality of the project is that the product that is suitable for the skin type of the user is recommended easily to the user based on real user reviews by just inputting a selfie of the user to the chatbot. It brings convenient and saves the time of the user.

REFERENCES

[1] N. J. Asmat Nizam Abdul-Talib, "Brand Consciousness and Brand loyalty: A study on Beauty Foreign Brand Beauty and Skin Care Products," 2020. [Online]. Available: https://www.igi-global.com/chapter/brand-consciousness-and-brand-loyalty/250543.

[2] S. Wolhuter, "The Beauty of the Bot," 2018. [Online]. Available: https://www.wearebrain.com/blog/innovation-and-transformation-strategy/chatbots-beauty-industry/.

[3] "We Tried The HelloAva Skincare App," [Online]. Available: https://ss.fintantaite.com/8434-we-tried-the-helloava-skincare-app-heres-our-thought.html.

[4] A. Carmn, "Sephora's latest app update," 2017. [Online]. Available: https://www.theverge.com/2017/3/16/14946086/sephora-virtual-assistant-ios-app- update-ar-makeup.

[5] "Virtual-Artist," [Online]. Available: https://www.sephora.my/pages/virtual-artist.

[6] "CoverGirl's Kalanibot Review," 2020. [Online]. Available: https://singaporechatbots.sg/covergirls-kalanibot-review/.

[7] J. P. a. O. Díaz, "A quality analysis of Facebook messenger's most popular chatbots," *Proceedings of the 33rd annual ACM symposium on applied computing,* no. pp. 2144- 2150, 10.1145/3167132.3167362, 2018.

[8] K. Johnson, "L'oréal beauty bot learning your style to make gift giving easier," 2017. [Online]. Available: https://venturebeat.com/2017/05/02/loreal-beauty-bot-learns-your- style-to-make-gift-giving-easier/.

[9] Peltarion, " Categorical Crossentrophy," 2021. [Online]. Available: https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy.

[10] R. Gomez, "Categorical Cross-Entropy Loss, Binary Cross-Entropy Loss, Softmax Loss, Logistic Loss, Focal Loss," 2018. [Online]. Available: https://gombru.github.io/2018/05/23/cross_entropy_loss/.

REFERENCES

[11] BotPenguin, "What Is A Telegram Bot? Reasons to use Bot for Telegram," 2021. [Online]. Available: https://chatbotslife.com/what-is-a-telegram-bot-reasons-to-use-bot-for-telegram-46b0d0579337#:~:text=A%20telegram%20bot%20can%20help,automate%20some%20of%20the%20interactions..

[12] N. Petherbridge, "RiveScript," 2022. [Online]. Available: https://www.rivescript.com/.

[13] KDnuggets, "Deep Learning Development with Google Colab, TensorFlow, Keras & Pytorch," 2018. [Online]. Available: https://www.kdnuggets.com/2018/02/google-colab-free-gpu-tutorial-tensorflow-keras-pytorch.html.

[14] J. Brownlee, "Introduction to the Python Deep Learning Library TensorFlow," 2016. [Online]. Available: https://machinelearningmastery.com/introduction-python-deep-learning-library-tensorflow/.

[15] J. A. K. M. M. S. S. W. Amin Ullah, "Action recognition in video sequences using deep bi-directional lstm with cnn features," no. IEEE Access, 6 , pp. 1155-1166, 2017.

[16] Tavish, "11 Important Model Evaluation Metrics for Machine Learning Everyone should know," 2019. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/.

APPENDIX

A - FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:1**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE No
2. WORK TO BE DONE Research on datasets Increase the number of datasets to improve the accuracy of the model
3. PROBLEMS ENCOUNTERED Find difficulty to find high quality datasets Data is collected manually and it is time consuming
4. SELF EVALUATION OF THE PROGRESS Need to work harder to collect high quality data

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

APPENDIX

(Project II)

Trimester, Year: Y2S3	**Study week no.:2**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE Collected datasets
2. WORK TO BE DONE Retrain and Retest Model with the datasets
3. PROBLEMS ENCOUNTERED Accuracy is still low after increasing the number of data
4. SELF EVALUATION OF THE PROGRESS Need to work hard to train the model

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	Study week no.:3
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE Model is trained with low accuracy
2. WORK TO BE DONE Retrain and retest model with multiple times and multiple parameters to reach the maximum score
3. PROBLEMS ENCOUNTERED No
4. SELF EVALUATION OF THE PROGRESS Need to work hard to retrain the model

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:4**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE

Model is trained with desirable accuracy

2. WORK TO BE DONE

Performance Evaluation

3. PROBLEMS ENCOUNTERED

4. SELF EVALUATION OF THE PROGRESS

Keep facing errors in model training, need improvement

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	Study week no.:5
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE
Model performance is evaluated
2. WORK TO BE DONE
Develop Chatbot
3. PROBLEMS ENCOUNTERED
No
4. SELF EVALUATION OF THE PROGRESS
Doing well

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:6**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE Halfway developing chatbot
2. WORK TO BE DONE Solve the bugs in chatbot development Testing Chatbot
3. PROBLEMS ENCOUNTERED Encountered bugs in developing chatbot
4. SELF EVALUATION OF THE PROGRESS Solve the bugs

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:7**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE Finish developing chatbot
2. WORK TO BE DONE Testing Chatbot
3. PROBLEMS ENCOUNTERED No
4. SELF EVALUATION OF THE PROGRESS Doing well

Supervisor's signature   Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:8**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE Finish Chatbot Testing
2. WORK TO BE DONE Combine Backend and Frontend Developments
3. PROBLEMS ENCOUNTERED No
4. SELF EVALUATION OF THE PROGRESS Satisfied for the progress I have

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:9**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE Combine Backend and Frontend Developments
2. WORK TO BE DONE A/B Testing
3. PROBLEMS ENCOUNTERED No
4. SELF EVALUATION OF THE PROGRESS Satisfied for the progress I have

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	**Study week no.:10**
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

1. WORK DONE A/B Testing
2. WORK TO BE DONE Write Project II Report Check everything
3. PROBLEMS ENCOUNTERED No
4. SELF EVALUATION OF THE PROGRESS Satisfied for the progress I have

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y2S3	Study week no.:11
Student Name & ID: Liew Yi Kei (18ACB02058)	
Supervisor: Dr. Aun Yichiet	
Project Title: Chatbot – Beauty Skin Care Products Recommendations	

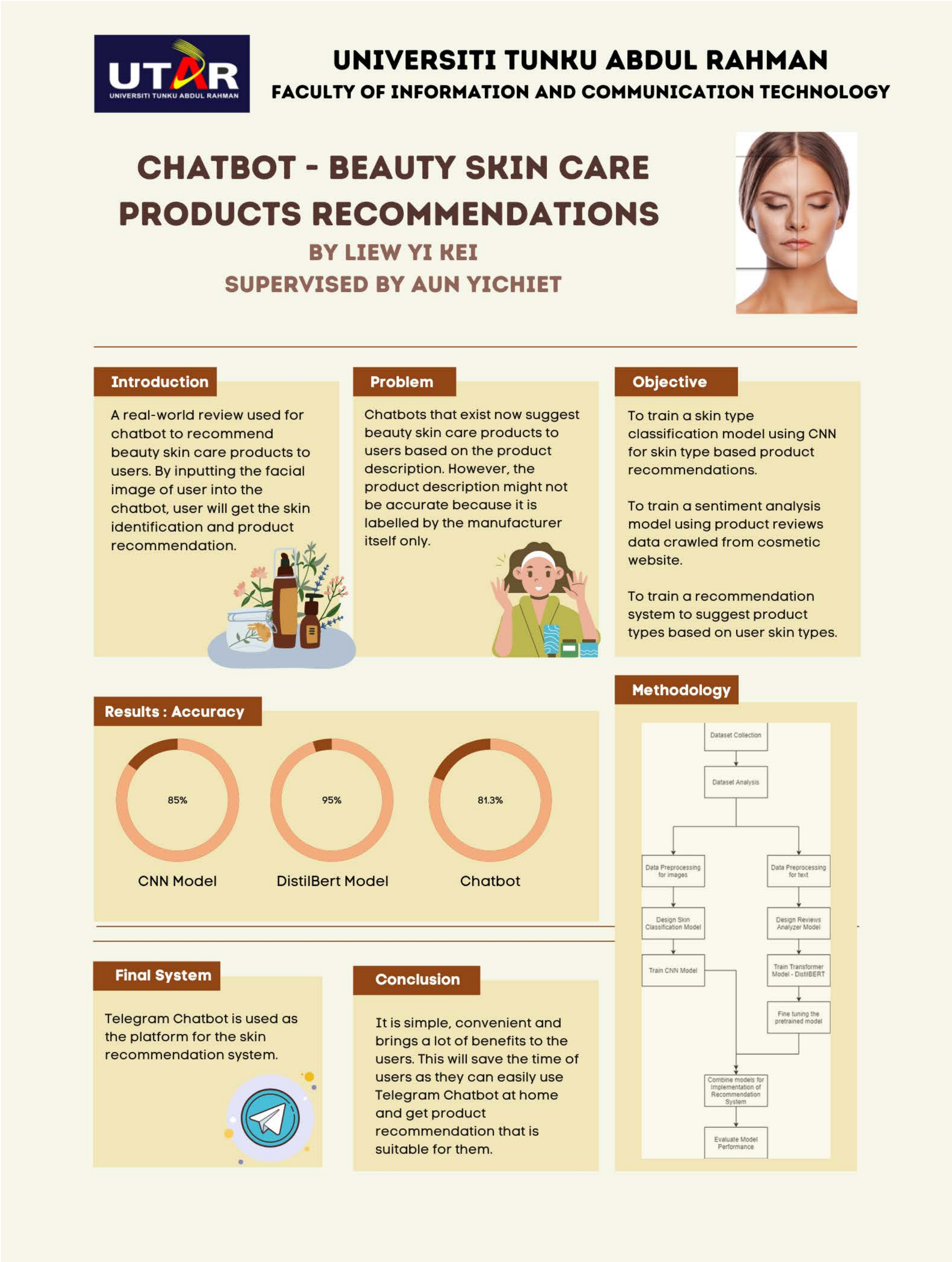
1. WORK DONE Finish Writing Project II Report
2. WORK TO BE DONE Check everything again
3. PROBLEMS ENCOUNTERED No
4. SELF EVALUATION OF THE PROGRESS Satisfied for the progress I have

Supervisor's signature

Student's signature

APPENDIX

B - POSTER



C - PLAGIARISM CHECK RESULT

27
CHAPTER 1 - INTRODUCTION

CHAPTER 1 – INTRODUCTION

1-1 Problem Statement and Motivation

Various chatbots have developed to recommend beauty skincare products to the customer. This project's **problem domain** is the chatbots that exist now suggest the products to users based on the product description. However, the product description might not be accurate because it is labelled by the manufacturer. The language on the ingredients list on the skincare products uses the International Nomenclature of Cosmetic Ingredients, so the users find it hard to understand what is written. The INCI has existed to create a standardized language of the ingredient names labelled on the product. However, it is not user friendly because some of the manufacturers attempt to appease the users by putting the more familiar name in parentheses next to the scientific name, for example, Alpha-tocopheryl acetate (vitamin E) to stop their complaints. The ingredients list will most likely look like a whole line of foreign language isolated by commas without this. Rather than doing investigator work on the ingredients list, chatbots are here to make our life easier by recommending products to users based on the product description labelled on the products. In this project, a **real-world review is used for chatbots** to recommend products to other related users.

1-2 Project Scope

Nowadays, the customer will look at the product description to determine if the product is suitable. This project develops an AI model to get the customers' actual feedback after using the product so that the users will get the most suitable products for them. When the products are recommended to the user, the AI model does not retrieve information from the product description stated from the product but from the users' reviews. For

APPENDIX

check turnitin

ORIGINALITY REPORT

15%	10%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	irep.iium.edu.my Internet Source	3%
2	www.wearebrain.com Internet Source	2%
3	Submitted to The Hong Kong Polytechnic University Student Paper	1%
4	www.igi-global.com Internet Source	1%
5	Submitted to Nanyang Technological University Student Paper	1%
6	venturebeat.com Internet Source	1%
7	Ruoyu Yang, Shubhendu Kumar Singh, Mostafa Tavakkoli, Nikta Amiri, Yongchao Yang, M. Amin Karami, Rahul Rai. "CNN-LSTM deep learning architecture for computer vision-based modal frequency detection",	1%

	Mechanical Systems and Signal Processing, 2020 Publication	
8	Submitted to Liverpool John Moores University Student Paper	1 %
9	Submitted to Babes-Bolyai University Student Paper	<1 %
10	www.analyticsvidhya.com Internet Source	<1 %
11	botfather.io Internet Source	<1 %
12	"Advances in Visual Informatics", Springer Science and Business Media LLC, 2019 Publication	<1 %
13	Submitted to Gitam University Student Paper	<1 %
14	Suriya Murugan, Chandran Venkatesan, M. G. Sumithra, Xiao-Zhi Gao, B. Elakkiya, M. Akila, S. Manoharan. "DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia From MR Images", IEEE Access, 2021 Publication	<1 %
15	Adem Tekerek. "A novel architecture for web-based attack detection using convolutional	<1 %

	neural network", Computers & Security, 2021 Publication		
16	ezug.msdve.it Internet Source	<1 %
17	www.thestar.com.my Internet Source	<1 %
18	machinelearningmastery.com Internet Source	<1 %
19	"Inventive Communication and Computational Technologies", Springer Science and Business Media LLC, 2021 Publication	<1 %
20	"Artificial Intelligence Applications and Innovations", Springer Science and Business Media LLC, 2020 Publication	<1 %
21	"Hybrid Intelligent Systems", Springer Science and Business Media LLC, 2021 Publication	<1 %
22	ebin.pub Internet Source	<1 %
23	norma.ncirl.ie Internet Source	<1 %
24	Submitted to Netaji Subhas Institute of Technology Student Paper	<1 %

25	Jamshid Mozafari, Afsaneh Fatemi, Parham Moradi. "A Method For Answer Selection Using DistilBERT And Important Words", 2020 6th International Conference on Web Research (ICWR), 2020 Publication	<1%
26	Submitted to University of Ghana Student Paper	<1%
27	documents.mx Internet Source	<1%
28	Banhi Sanyal, Ramesh Kumar Mohapatra, Ratnakar Dash. "Real-time Embedded TSR using MDEffNet: An efficient compact DNN on varied multi-resolution images across multiple databases", International Journal of Wavelets, Multiresolution and Information Processing, 2020 Publication	<1%
29	Submitted to Higher Education Commission Pakistan Student Paper	<1%
30	punchng.com Internet Source	<1%
31	Submitted to Universiti Teknologi Petronas Student Paper	<1%
32	www.synthedit.com Internet Source	

			<1 %
33	De Simone, Francesca, Lutz Goldmann, Jong-Seok Lee, and Touradj Ebrahimi. "", Applications of Digital Image Processing XXXIV, 2011. Publication	<1 %
34	Submitted to University of Wales central institutions Student Paper	<1 %
35	tdn.com Internet Source	<1 %
36	Chin Kit Ng, Soon Nyean Cheong, Wen Wen-Jiun Yap, Yee Loo Foo. "Outdoor Illegal Parking Detection System Using Convolutional Neural Network on Raspberry Pi", International Journal of Engineering & Technology, 2018 Publication	<1 %
37	toptenproducts.com Internet Source	<1 %

APPENDIX

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.:

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Liew Yi Kei
ID Number(s)	18ACB02058
Programme / Course	Business Information Systems
Title of Final Year Project	Chatbot – Beauty Skin Care Products Recommendations

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits**
Overall similarity index: 15 **%** **Similarity by source** Internet Sources: 10 % Publications: 7 % Student Papers: 4 %	
Number of individual sources listed of more than 3% similarity: 0	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor	Signature of Co-Supervisor
Name: Dr. Aun YiChiet	Name: ______________________
Date: 21/4/22	Date: ______________________

D - PROJECT 2 CHECKLIST

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	18ACB02058
Student Name	Liew Yi Kei
Supervisor Name	Dr. Aun YiChiet

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
	Front Plastic Cover (for hardcopy)
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
√	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
√	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

APPENDIX

I, the author, have checked and confirmed all the items listed in the table are included in my report. ________________________ (Signature of Student) Date: 21/4/22