AI-Powered

Do-It-Yourself (DIY)

Diabetes HealthApp

Team Members

Sim Yuh Fan

Sahendra Pang

Santi

Tan Xiang Feng

Zhang Zhewei

CONTENTS

1 INTRODUCTION

Diabetes is a worldwide health problem that have affected 400 million people worldwide. There are two major types of diabetes:

- Type 1 – often diagnosed in childhood
- Type 2 – largely caused by an unhealthy lifestyle and diet

From a study by International Diabetes Federation (IDF) done in 2015, it revealed that Singapore had the second highest proportion of diabetes among developed nations. The number of diabetic cases were projected to further increase by another 200K to 600K by 2030. In terms of long-term complications, we had the highest rate of lower limb amputation in the world with about 1,200 diabetes patients undergoing amputation every year.

While there is no cure for diabetes, Type 2 diabetes can be managed and prevented through one's lifestyle. By leveraging on the power of machine learning and Artificial Intelligence ('AI'), we aim to develop a tool to

1) Allow individual to self-assess their risk of having type 2 diabetes
2) Help individual manage diabetes
3) Look out for signs of complications such as blindness and foot problems (in the next release).

This is achieved using an AI-powered Do-It-Yourself (DIY) Diabetes HealthApp (or in short "Diabetes HealthApp") as a "minimum viable product" (MVP). The first release can be improved upon successive releases with more user experience.

2 BUSINESS CASE

2.1 Problem Statement

While Singapore has one of the highest proportions of diabetes cases, the Ministry of Health believes that a third of the diabetes patient are unaware that they are ill (NRDO).

This is exacerbated by barriers to health screenings such as not knowing where to get free screening, procrastination, and not seeing the point of going for screening without subsequent follow up or explanation from a healthcare provider, as noted by a nationwide study examining the knowledge, attitudes and practices (KAP) pertaining to diabetes in Singapore (IMH).

However, with early detection, one can manage and prevent diabetes. If left untreated, the situation may worsen, leading to long term complications such as blindness and lower limb amputation.

2.2 Proposed Solution

We have developed a phone app covering the 3 functions (aka Strategy) stated in the Section 1. Individual will be able to predict their risk of having type 2 diabetes through some simple details such as height and weight. Individual can also detect Diabetic Retinopathy by simply uploading a photo of their retina to the app. Lastly, by filling up a questionnaire on their dietary intake and physical activities, we will be able to recommend if one is doing enough to manage diabetes.

This solution is able to tackle the above problem statement as it is:

1. Convenience – One can self-assess at their own convenience or at the comfort of their home.
2. Zero Cost – For the low-income/ less privileged segment, they can get themselves regularly checked for free.

This greatly lower the barriers to health screenings. The high penetration rate of mobile devices in Singapore also allows this solution to be widely available to the entire population. Furthermore, this reduces the strain on healthcare system as individuals can get themselves “checked” using the app prior to visiting the doctor if necessary.

2.3 Market Research

Diabetes is a worldwide health issue hence it is not surprising to see a number of products available in the market that provide medical diagnosis for diabetes. One notable mention will be Singapore Eye LEson Analyzer+ (SELENA+) developed by National Eye Centre. SELENA employs the latest image analysis and state-of-the-art machine learning techniques to serve as an automated, real-time detection tool that can match human grading (Aisyah). It has been implemented at all polyclinics and some 20 optometry stores.

However, barriers to screening still exist for SELENA, as individuals are required to be physically present. Our solution is able to lower the barriers to screening, by bringing convenience to one’s doorstep, simply with the click of a phone app.

3 PROJECT OVERVIEW

Full Name	Student ID	Roles and Responsibility
Sim Yuh Fan	A0249251E	• Administrative Lead • UI/UX Development • Backend Application Development
Sahendra Pang	A0135877N	• Backend Application Development • Model Development and Testing (Strategy 3) • Video
Santi-	A0249294R	• Application Design and Development • Model Development and Testing (Strategy 1 & 2) • Project Report Writing
Tan Xiang Feng	A0072452L	• Model Development and Testing (Strategy 3) • Project Report Writing • Video
Zhang Zhewei	A0249258R	• Model Development and Testing (Strategy 1) • Project Report Writing • Video

4 PRODUCT DESIGN

4.1 Main Features

To combat diabetes and at the same time bringing convenience to everyone, we have developed a phone app with 3 functionalities (aka Strategy):

- **Strategy 1: Predicting Risk of Diabetes**

  Individual will be able to predict their risk of having type 2 diabetes through some simple details such as height and weight.

- **Strategy 2: Lifestyle Recommender**

  Individuals might be clueless in the effectiveness of their physical activities and dietary plan in reducing the risk of diabetes. Leveraging on our reasoning system, one will have a better idea if their dietary intake and physical activities enough to reduce risk of diabetes.

- **Strategy 3: Detect Diabetic Retinopathy**

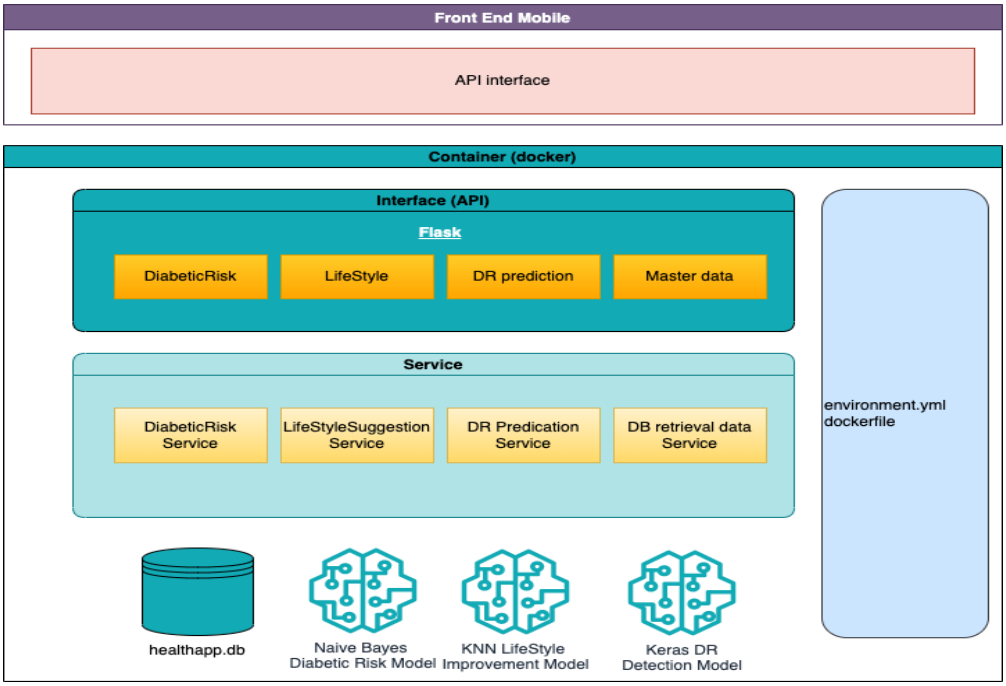
  Individual will be to detect Diabetic Retinopathy by simply uploading a photo of their retina to the app. By doing so, diabetic patient will be able to keep themselves “checked” and reduce the frequent visit to the doctor.

4.2 System Design

The diagram below provides us a broad overview of the architecture. It has 2 major components

1. App - Built with Android Studio and has an intuitive user interface.
2. Engine – Built with python and deployed in docker container. It supports Flask API services, Scikit-learn & Keras and Sqlite database.

Figure 1: Overall Architecture



5 SYSTEM MODELLING

For the first release of the AI-powered Do-It-Yourself (DIY) Diabetes HealthApp, we will be covering the 3 Strategy stated in Section 4.1.

5.1 Strategy 1: Predicting Risk of Diabetes

5.1.1 Data Preparation

For Strategy 1, our objective is to help individual predict the risk of diabetes through machine learning model. For this strategy, we will be using 2013 and 2015 data from Behavioural Risk Factor Surveillance System (BRFSS). BRFSS is U.S. nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviours, chronic health conditions, and use of preventive services. It was established in 1984 and till date they have collected in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.

There is a total of 330 features in BRFSS dataset. The project team have vetted through the long list and have picked the 15 features that we think are more relevant in predicting diabetes.

The list of features is as per below.

Table 5-1: Summary of Features

No.	Feature Name	Feature Description	Values
1	HighBP	High Blood Pressure	0: no high blood pressure 1: have high blood pressure
2	HighChol	High Cholesterol	0: no high cholesterol 1: have high cholesterol
3	BMI	Body Mass Index	
4	Smoker	Smoker	0: otherwise 1: have smoked at least 100 cigarettes in your entire life
5	Stroke	Stroke	0: no stroke 1: have incident of stroke
6	HeartDiseaseorAttack	Have Coronary Heart Disease (CHD) or Myocardial Infraction (MI)	0: no 1: yes
7	PhysActivity	Have some form of physical activity in the past 30 days	0: no 1: yes
8	Fruits	Have consume ≥ 1 serving of fruits per day	0: no 1: yes
9	Veggies	Have consume ≥ 1 serving of vegetables per day	0: no 1: yes

No.	Feature Name	Feature Description	Values
10	HvyAlcoholConsump	Male: Have consume >14 portion of drinks per week Female: Have consume >7 portion of drinks per week	0: no 1: yes
11	MentHlth	For the past 30 days, how many days are you feeling stress, depression or having negative emotions	
12	PhysHlth	For the past 30 days, how many days you have physical illness and injury	
13	DiffWalk	Do you have seriour difficulty walking or climbing stairs	0: no 1: yes
14	Sex	Gender	0: female 1: male
15	Age	Age	

5.1.2 Algorithm Choice and Model Testing

Different algorithms, from the simpler ones like Naïve Bayes to the more sophisticated algorithm like Random Forest (RF), were tested to search for one that is both powerful and easily explainable. Generally, there is a trade-off between accuracy and model interpretability, which means the better the model output becomes, the more complex the algorithm is. Nonetheless, it is essential to try simple algorithms in addition to complex ones as simple algorithm can provide a benchmark performance. Moreover, if the simple algorithm can achieve acceptable accuracy, it will hold advance over more complex ones because it is highly interpretable.

Below is the high-level tabulation in terms of accuracy of the considered algorithms.

Table 5-2: Results of Various Algorithms

Algorithm	Training	Validation
Naive Bayes	**68%**	**69%**
Decision Tree	92%	65%
k-Nearest Neighbour	N.A	~ 69%
Random Forest	N.A	~45% (based on 50, 100, 150, 200, 300, 500, 1000 estimators)

For the given dataset, based on the above results, we have decided to use Naïve Bayes Model as it is easier to explain and more stable across time; smaller gap between the training and validation dataset indicates the model is more stable with less over-fitting.

5.2 Strategy 2: Lifestyle Recommender

5.2.1 Data Preparation

While there is no cure for diabetes, Type 2 diabetes can be managed and prevented through one's lifestyle. Leveraging on our reasoning system (combination of ML model and rule-based engine), one will have a better idea if their dietary intake and physical activities enough to reduce risk of diabetes.

For the ML model, we will be using data from NHANES to project the BMI of the individual based on their dietary intake and physical activities. The list of features is as per below.

Table 5-3: Summary of Demographic Features

No	Feature Name	Feature Description	Values
1	id	Unique numeric identifier of the participant	
2	sex	Sex of the participant	1: Male 2: Female
3	age	Age of the participant in years	
4	wgt	Body weight of the participant in kg	
5	hgt	Height of the participant in cm	
6	pa	Physical activity level of the participant in Metabolic Equivalents of Task (METs) minutes per week	
7	special_diet	Particular dietary pattern that the participant followed at the time of the dietary data collection	0: No special diet 1: Vegetarian diet 2: Vegan diet 3: Weight loss diet 4: Weight gain diet 5: Other
8	clin_diet	Particular dietary pattern that the participant followed at the time of the dietary data collection due to a health condition	0: No clinical diet 1: Low fat or cholesterol diet 2: Low salt or sodium diet

No	Feature Name	Feature Description	Values
			3: Diabetic diet 4: Renal or kidney diet 5: Gluten-free or celiac diet 6: Other

Table 5-4: Summary of Dietary Features

No	Feature Name	Feature Description	Values
1	meal_type	Type of the consumed meal	1: Before breakfast 2: Breakfast 3: Snack between breakfast and lunch 4: Lunch 5: Snack between lunch and dinner 6: Dinner 7: Snack after dinner 8: Brunch 9: Snack (unspecified when) 10: Drink 11: Feeding (infant only) 12: Extended consumption 13: Merienda
2	energy	Energy intake from simple food/ingredient	
3	totalpro	Total protein intake from simple food/ingredient	

No	Feature Name	Feature Description	Values
4	carb	Total carbohydrate intake from simple food/ingredient	
5	fiber	Total dietary fiber intake from simple food/ingredient (fruits, vegetables, grains, legumes, pulses), defined as the carbohydrate polymers which are not hydrolyzed by the endogenous enzymes in the small intestine of human beings. Dietary fiber should optimally be quantified using the AOAC method of analysis	
6	mufa	Total monounsaturated fat intake from simple food/ingredient	
7	chol	Cholesterol intake from simple food/ingredient	
8	pufa	Total poly-unsaturated fat intake from simple food/ingredient	
9	totalfat	Total fat intake from simple food/ingredient	
10	sfa	Total saturated fat intake from simple food/ingredient	
11	water	Total water intake from simple food/ingredient	
12	totalsugars	Total sugar intake from simple food/ingredient	

Below outlines the steps taken to obtain the final training model. For more detailed information, you can refer to 'DataPrep_LS.ipynb'

1. Merged participant data and diet data based on participant id
2. Remove duplicate value
3. Remove data with all value empty
4. Calculate bmi with formula bmi = wgt / (hgt/100)^2
5. Sum water intake
6. Categorize mealtime
7. Sum all value based on category of meal

5.2.2 Algorithm Choice and Model Testing

We have tested the dataset with k-Nearest Neighbor and Decision Tree Algorithm. From the results below, we have chosen k-NN as the model to predict the BMI.

Table 5-5: Results of Hyperparameter Tuning (k-NN Model)

k-NN	Train score	Test Score
1	0.999992	1
2	0.999992	1
3	**0.999992**	**1**
4	0.999962	0.999989
5	0.999902	0.99996
6	0.999797	0.999907
7	0.999623	0.999813
8	0.999363	0.999664
9	0.998995	0.99944
10	0.998497	0.999119
11	0.997854	0.998684
12	0.997035	0.998118
13	0.996011	0.997391
14	0.994779	0.996487
15	0.993342	0.995399
16	0.991667	0.994102

Table 5-6: Results of Hyperparameter Tuning (Decision Tree)

Depth	Accuracy
1	0.078318
2	0.078318
3	0.079579
4	0.087404
5	0.092932
6	0.103229
7	0.116853
8	0.135912
9	0.158935
10	0.188622
11	0.228308
12	0.272019
13	0.321484
14	**0.381219**
15	0.439808
16	0.493334

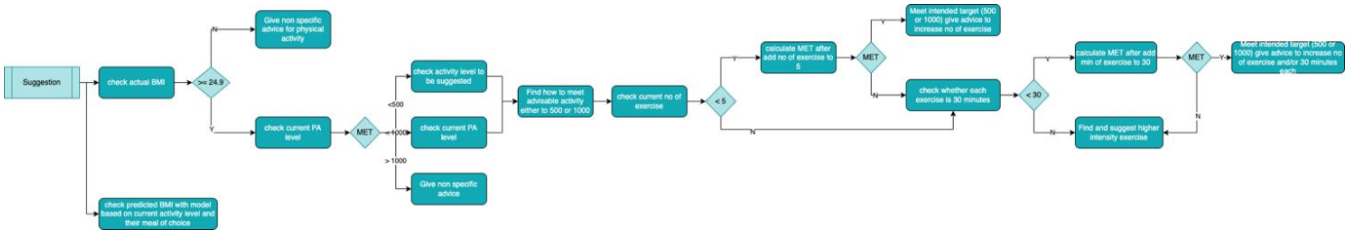
The output from the above model will be passed into the below rule-based engine. Based on some simple rules, it will suggest if the current dietary intake and physical activities enough to reduce risk of diabetes.

Table 5-7: Summary of the Rule-based Engine

No	Input	Usage	Recommender
1	Age	Use for model	
2	Weight	Calculate BMI BMI = W / H^2	
3.	Height		
4	Physical activity of choice		Physical activity recommendation

No	Input	Usage	Recommender
5	Level of activity	System will use this to calculate the ideal physical activities and relate this to current BMI. Target of ideal activities is between 500 to 1000 MET	
6	No of exercise a week		
7	Duration of each exercise		
8	Breakfast	Pass the information of recipe detail into LifeStyle model to get the predicted bmi output.	Advice to get healthier breakfast if model return bmi > 24.9
9	Lunch		Advice to get healthier lunch if model return bmi > 24.9
10	Dinner		Advice to get healthier dinner if model return bmi > 24.9

Figure 2: Pictorial View of the Rule-based Engine



5.3 Strategy 3: Detect Diabetic Retinopathy

5.3.1 Data Preparation

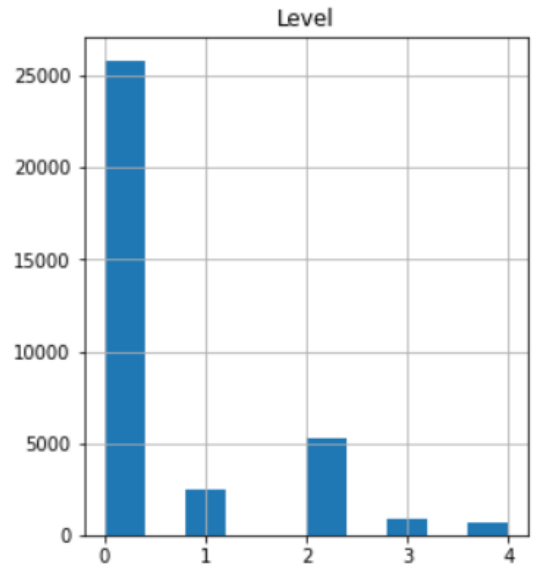
For Strategy 3, our objective is to help individuals detect diabetic retinopathy. For the model, we will be using data from Kaggle. There is at total of 37,070 high-resolution images. A clinician has rated the presence of Diabetic Retinopathy ('DR') in each image on a scale of 0 to 4, according to the following scale:

Table 5-8: Level of Diabetic Retinopathy

Scale	Description
0	No DR
1	Mild DR
2	Moderate DR
3	Severe DR
4	Proliferative DR

As part of the explanatory data analysis, we noticed that the data is imbalanced with 70% of the subject having no DR. Building a model on imbalanced dataset may lead to a sub-optimal model with seemingly high accuracy. To tackle the imbalanced dataset, we have performed stratified sampling, picking 200 subjects for each subgroup (i.e., different levels and different eyes). We have split the 2,000 subjects (i.e. 200 x 5 levels x 2 eyes) into training and validation dataset with 60:40 split.

Table 5-9: Frequency by the Level of Diabetic Retinopathy



	Level	LeftEye	count
0	0	0	12939
1	0	1	12871
2	1	0	1231
3	1	1	1212
4	2	0	2590
5	2	1	2702
6	3	0	448
7	3	1	425
8	4	0	355
9	4	1	353

Before training the model, we have performed some pre-processing to resize the image to 224 x 224 and to standardize the data to ensure that the RGB channel is in the [0, 1] range.

5.3.2 Algorithm Choice and Model Testing

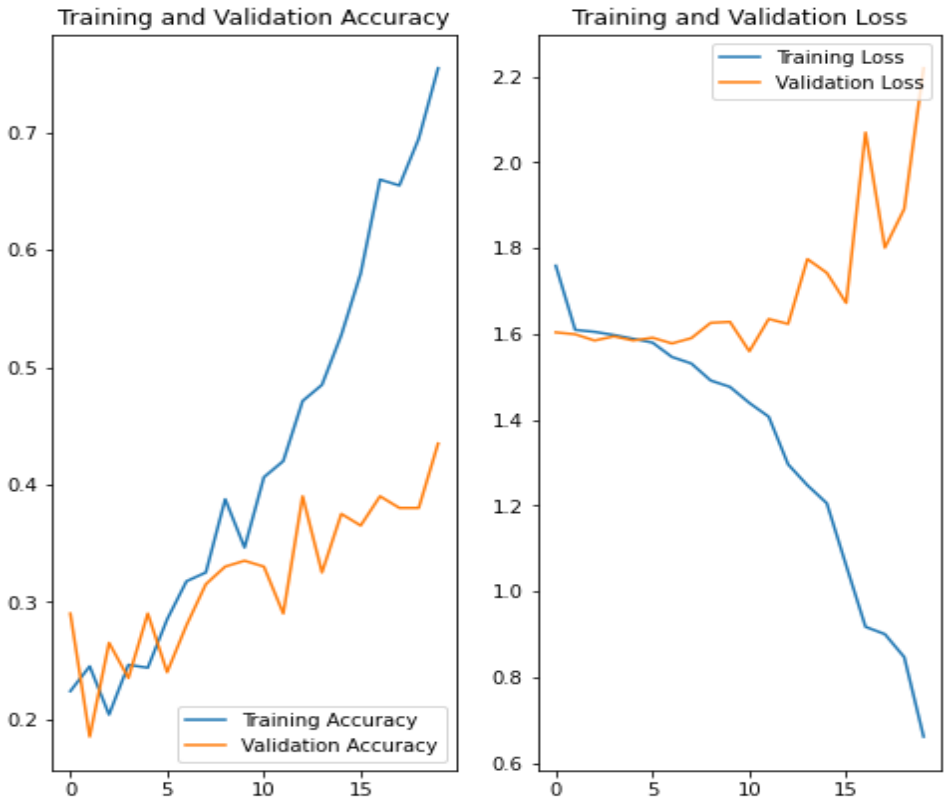
For this strategy, we will be using one of the more popular deep learning frameworks – TensorFlow. We first define the convolutional base using stack of Conv2D and MaxPooling2D layers. We then activate it using LeakyReLu activation function. Lastly to complete the model, we have flattened the last output tensor from the convolutional base into 2 Dense layer. For this strategy, there is 5 output classes, so we have used a final Dense layer with 5 outputs.

Table 5-10: Convolutional Neural Network

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 224, 224, 32)      896
_________________________________________________________________
leaky_re_lu (LeakyReLU)      (None, 224, 224, 32)      0
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 112, 112, 32)      0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 112, 112, 64)      18496
_________________________________________________________________
leaky_re_lu_1 (LeakyReLU)    (None, 112, 112, 64)      0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 56, 56, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 56, 56, 128)       73856
_________________________________________________________________
leaky_re_lu_2 (LeakyReLU)    (None, 56, 56, 128)       0
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 28, 28, 128)       0
_________________________________________________________________
flatten (Flatten)            (None, 100352)            0
_________________________________________________________________
dense (Dense)                (None, 128)               12845184
_________________________________________________________________
leaky_re_lu_3 (LeakyReLU)    (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 5)                 645
=================================================================
Total params: 12,939,077
Trainable params: 12,939,077
Non-trainable params: 0
_________________________________________________________________
```

We then compile and train the model using Adam optimizer with loss function “categorical crossentropy”. From the results shown below, we noted that divergence in the loss function between the training and validation set. This indicates overfitting in the training dataset.

Figure 3: Model Performance



To prevent overfitting, we have repeated the above Sequential Model with the addition of 2 Dropout Layer. The difference between the loss function has dropped (i.e. Training: 1.4, Validation: 1.7).

Table 5-11: Convolutional Neural Network (with Dropout Layer)

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_3 (Conv2D)            (None, 224, 224, 32)      896
_________________________________________________________________
leaky_re_lu_4 (LeakyReLU)    (None, 224, 224, 32)      0
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 112, 112, 32)      0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 112, 112, 64)      18496
_________________________________________________________________
leaky_re_lu_5 (LeakyReLU)    (None, 112, 112, 64)      0
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 56, 56, 64)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 56, 56, 128)       73856
_________________________________________________________________
leaky_re_lu_6 (LeakyReLU)    (None, 56, 56, 128)       0
```

```
_________________________________________________________________
dropout (Dropout)            (None, 56, 56, 128)       0
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 28, 28, 128)       0
_________________________________________________________________
flatten_1 (Flatten)          (None, 100352)            0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               12845184
_________________________________________________________________
leaky_re_lu_7 (LeakyReLU)    (None, 128)               0
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_3 (Dense)              (None, 5)                 645
=================================================================
Total params: 12,939,077
Trainable params: 12,939,077
Non-trainable params: 0
```

The model is satisfactory with an accuracy of 42% for the training dataset and 32% for the validation dataset. With that, we have implemented this model for our release 1. As for the next step, to improve model accuracy, we will be trying augmentation to increase the dataset and trying pre-trained model such as VGG Model to increase the accuracy.

6 SYSTEM DEVELOPMENT & IMPLEMENTATION

We have developed a phone app using Android Studio, hoping to combat diabetes and at the same time bringing convenience to everyone. Below shows the UI of our Diabetes HealthApp with the 3 features for user to chose from.

Table 6-1: User Interface (Main Page)

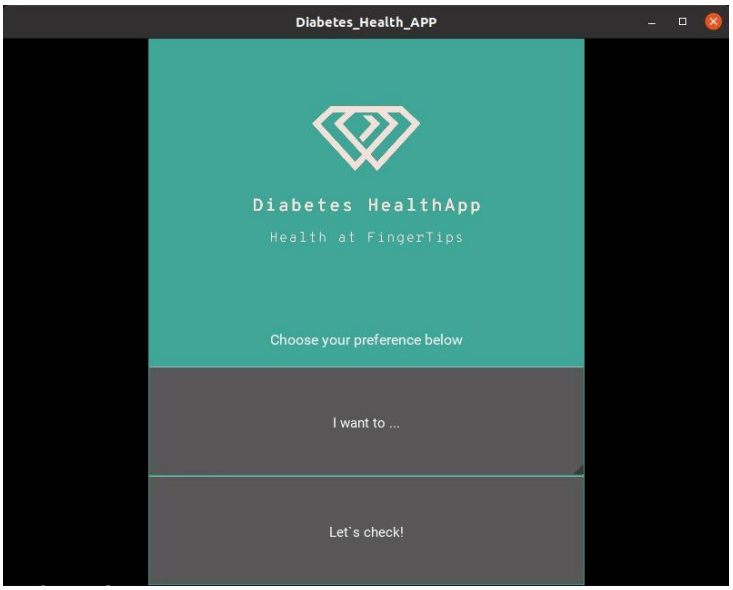
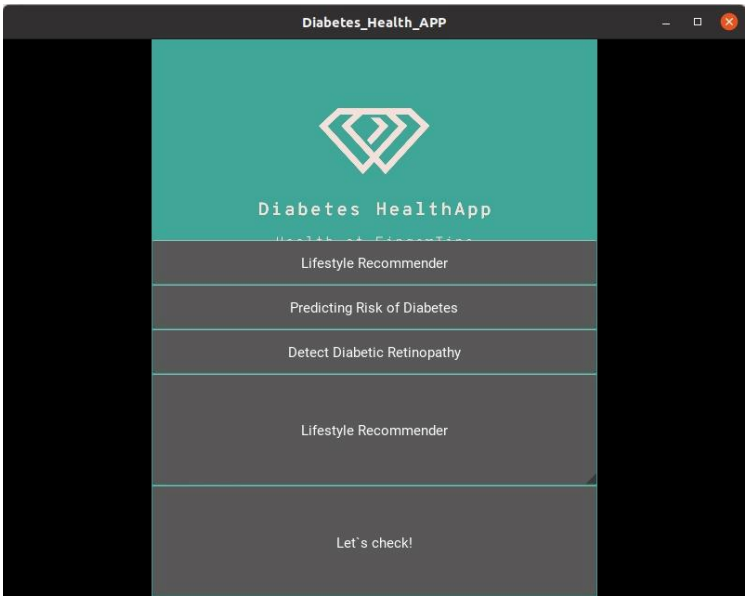


Table 6-2: User Interface (Features for user to select)



The below diagram shows the connectivity between the App and the backend module for the 3 features. To see how it looks like on the app, please refer to user manual.

Figure 4: Process flow for Strategy 1 - Predicting Risk of Diabetes

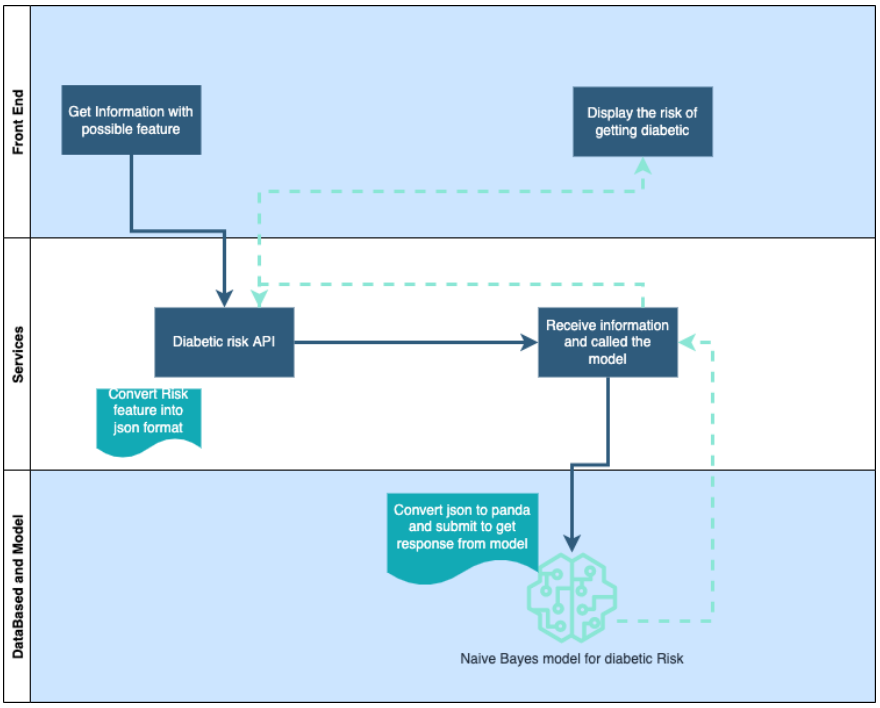


Figure 5: Process flow for Strategy 2 - Lifestyle Recommender

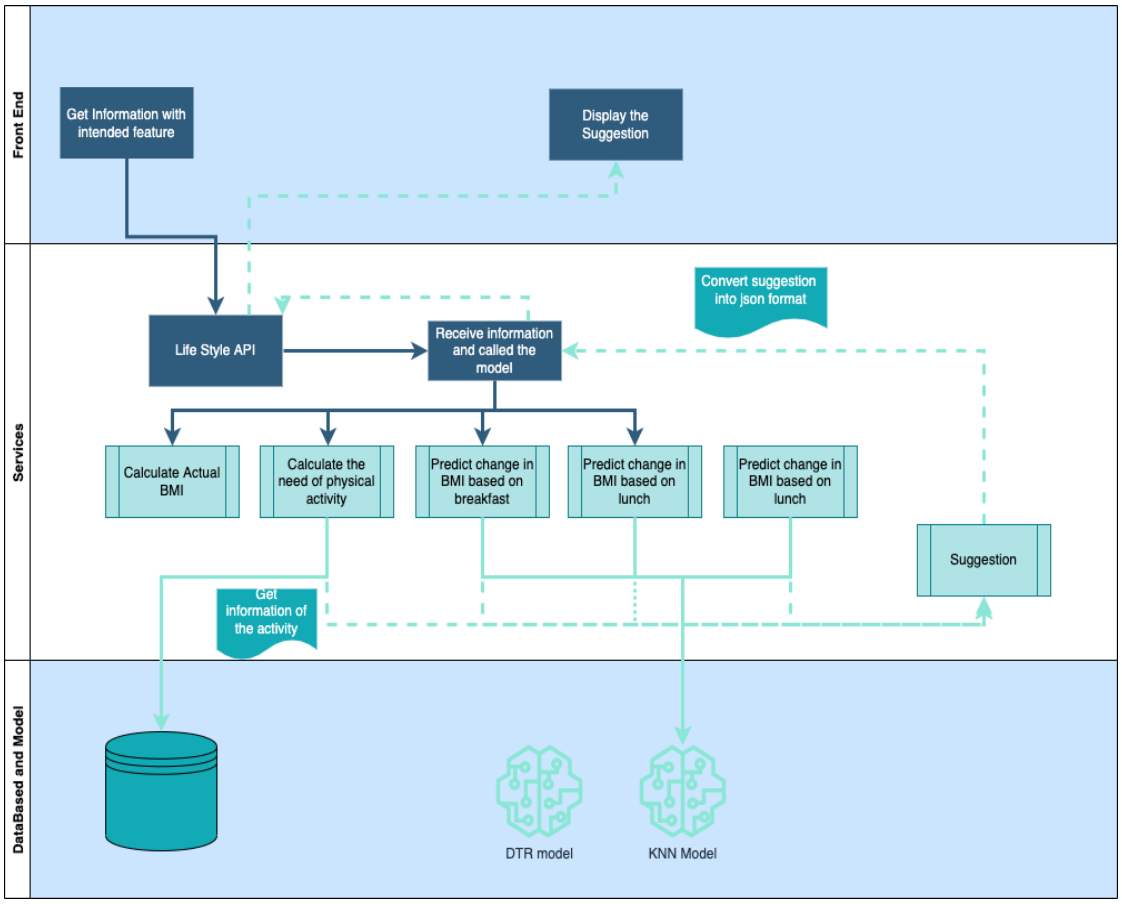
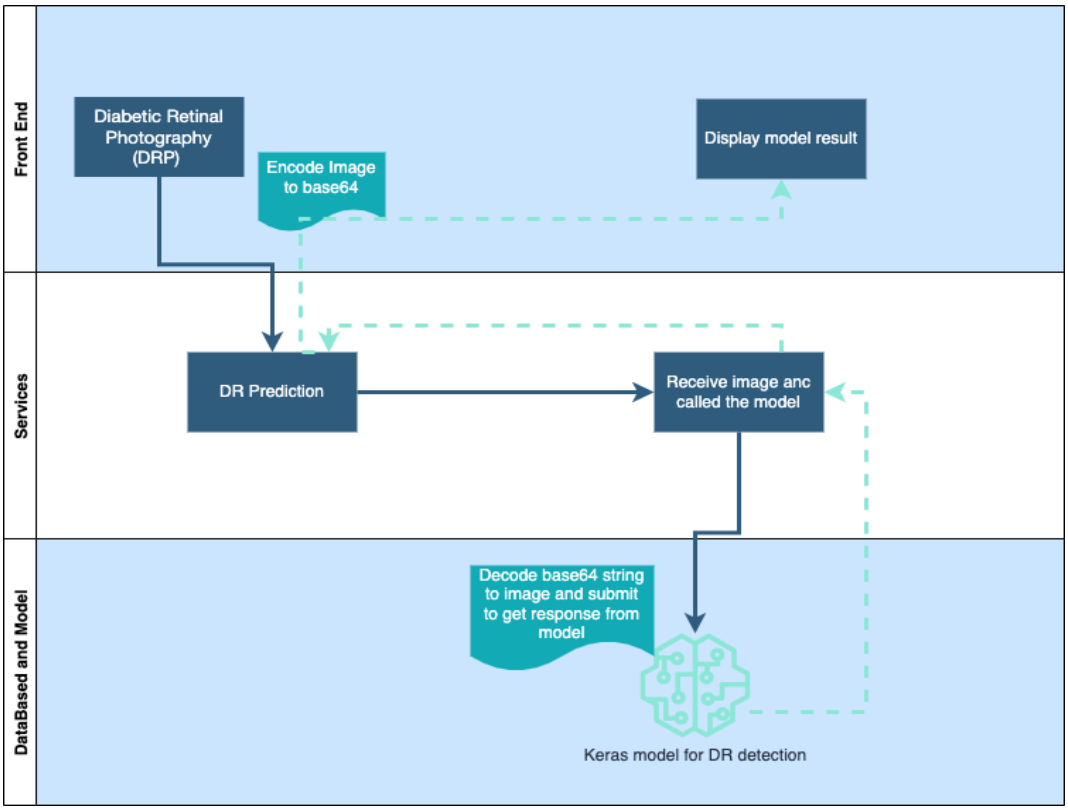


Figure 6: Process flow for Strategy 3 - Detect Diabetic Retinopathy



7 CONCLUSION

We have successfully implemented a phone app that allow individual to predict risk of diabetes, to better manage diabetes and to detect diabetic retinopathy at the comfort of their home. As for the low-income/ less privileged segment, they can now get themselves checked regularly for free.

7.1 Future Development

This phone app serves as an MVP, and it will be improved upon in successive releases with more use experience. This product has the potential to

1) Cover the full life cycle of diabetes
2) Connect with the doctor via the app
3) Build new verticals to detect health problems (i.e., Heart Diseases, Diabetes, Thyroid disorder, Cancer, and Multiple sclerosis) though our AI-enabled scanner.

8 REFERENCES & DATASETS

Aisyah, Kirana. [Online] https://opengovasia.com/singapores-use-of-ai-in-healthcare-and-education/.

IMH. [Online] https://www.imh.com.sg/Documents/research-announcements/20Apr2021_Singapore%20residents%20show%20a%20high%20recognition%20of%20diabetes.pdf.

NRDO. [Online] https://www.nrdo.gov.sg/docs/librariesprovider3/default-document-library/diabetes-info-paper-v6.pdf?sfvrsn=0..

Strategy 1: Predicting Risk of Diabetes

https://www.kaggle.com/alexteboul/diabetes-health-indicators-dataset

Strategy 2: Lifestyle Recommender

National Health and Nutrition Examination Survey Data (NHANES), 2015-2016; Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Harmonized for the Global Dietary Database. Accessed at www.globaldietarydatabase.org/management/microdata-surveys

Strategy 3: Detect Diabetic Retinopathy

https://www.kaggle.com/c/diabetic-retinopathy-detection/

9 APPENDIX

9.1 Appendix A – Project Proposal

Date of proposal:

28 Feb 2022

Project Title:

ISS Project – Diabetes-Health-App

Group ID (As Enrolled in LumiNUS Class Groups):

Group Members (Name, Student ID):

Sim Yuh Fan	A0249251E
Sahendra Pang	A0135877N
Santi-	A0249294R
Tan Xiang Feng	A0072452L
Zhang Zhewei	A0249258R

Sponsor/Client: *(Name, Address, Telephone No. and Contact Name)*

None

Background/Aims/Objectives:

<u>Background</u>

Diabetes is a worldwide health problem that have affected 400 million people worldwide. While Singapore has one of the highest proportions of diabetes cases, the Ministry of Health believes that a third of the diabetes patient are unaware that they are ill.

This is exacerbated by barriers to health screenings such as not knowing where to get free screening, procrastination, and not seeing the point of going for screening without subsequent follow up or explanation from a healthcare provider.

<u>Aim/ Objective</u>

Leveraging on the power of machine learning and Artificial Intelligence ('AI'), we aim to develop a tool to

1) Allow individual to self-assess their risk of having type 2 diabetes
2) Help individual manage diabetes

3) Look out for signs of complications such as blindness and foot problems

in which individual can self-assess at their own convenience at zero cost.

Requirements Overview:

- Research ability
- Programming ability
- System integration ability

Project Descriptions:

To combat diabetes and at the same time bringing convenience to everyone, we have developed a phone app with 3 functionalities (aka Strategy):

- **Strategy 1: Predicting Risk of Diabetes**

  Individual will be able to predict their risk of having type 2 diabetes through some simple details such as height and weight.

- **Strategy 2: Lifestyle Recommender**

  Individuals might be clueless in the effectiveness of their physical activities and dietary plan in reducing the risk of diabetes. Leveraging on our reasoning system, one will have a better idea if their dietary intake and physical activities enough to reduce risk of diabetes.

- **Strategy 3: Detect Diabetic Retinopathy**

  Individual will be to detect Diabetic Retinopathy by simply uploading a photo of their retina to the app. By doing so, diabetic patient will be able to keep themselves “checked” and reduce the frequent visit to the doctor.

9.2 Appendix B – Mapped System Functionalities against knowledge, techniques, and skills of module courses.

Module Courses	System Functionalities/ Techniques Applied
Machine Reasoning (MR)	Knowledge Representation Supervised Learning Algorithms (k-NN, Decision Tree, Naïve Bayes) Rule Based System
Reasoning Systems (RS)	Retrieval Reasoning Systems (Analytic Task)
Cognitive System (CGS)	Vision Cognitive Semantic Network

9.3 Appendix C – Individual Report

9.3.1 Personal Report: Sim Yuh Fan

Personal Contribution

Administrative Lead

- Organize and host team discussion
- Communicate and share information with the lecturers and classmates.

UI/UX Development

- Apply graphical design skill to enhance user experience e.g. simplification of user interface content, implementation of tools (toggle switch, drop-down list, image-viewer).

Backend Application Development

- Apply different types of Machine Learning techniques on real diabetes dataset.
- Learn how to build an Android application from nothing to everything.
- Use Python tool to design an Android application which is capable to communicate with online server.
- Utilize Android platform to present Machine Learning model to general public.

Lessons Learnt

From this hands on project, I have learnt a lot of things from my team and learn by doing. Below are the significant learning outcomes:

- Improve a lot in coding skills
- Learn how to use Python to build the android application
- Gain knowledge about basic Python language and use concepts such as variables, and conditional
- Gain the opportunity to work with others as a team towards a goal

Future

With the knowledge and skills gain from this hands on project, I able to use these to solve real-life problems. This is because I understand the mechanism behind few machine learning techniques and will apply the skills to overcome the issues in future.

9.3.2 Personal Report: Sahendra Pang

Personal Contribution

Retinopathy Model

- Data prepossessing: Filter out black images, rotate, mirror images, dataset balancing
- Build model: Study CNN, build, train and testing (optimes hyper parameter)

Android application

- Testing on Kivy platform
- Testing on APK generation using buildozer
- Testing on API call

Documentation

- Technical Presentation video

Lessons Learnt

Throughout this project I have a lot of things to learn from my team mate as well as learning myself by doing. Following are the most significant learning:

- Improve a lot in python programming
- Data preprocessing is very important as unbalance dataset may result to low accuracy in production
- As this is the first time, I am doing AI solution implementation, I experience the process such as gathering data, data clean up, model building, training and testing the model, transforming the result to UI.
- Compiling python into android application
- Team work, collaboration and time management.

Future

With the knowledge gain through out this project and in the class, I can try few machine learning algorithms that is suitable for the problem in my workplace. I could also implement the same process in implementing AI solution such as data gathering, cleaning of the data, model building / selection and last but not least the data visualization.

9.3.3 Personal Report: Santi

Personal Contribution

Based on my personal experience in application development, my main role is to help with the system design and development of the engines, such as connectors and services to the model and database. I also build the reasoning features.

I also help the team to put up all engine deployment, as well as to containerize the services to ease future installation as from machine to machine some versions that we used might not be compatible with the operating systems.

I also work with other team member to find data source, do data cleanup, and develop the model that being used by our application. In term of report and videos I also worked with the team to produce the system design and some other area.

Lessons Learnt

Throughout this project in term of technicality I learned things that I'm not familiar with before such as

- How to clean data to become a meaningful source of model, this apparently took us quite sometimes
- How to write and develop a running client/server application in python (my background mainly is Java)
- How to deploy the application in docker. This I found it's very meaningful for python as I did realize some of the components that we are using are very sensitive towards operating systems that we use.

I also learnt other things from my team member as part of knowledge that not related to software development such as:

- During the model creation, there are many things that I learned in the class come handy and often my project team will help if there are phenomenon that happened to the model that I don't really understand.
- There are many aspect to consider to get a meaningful AI model

Nevertheless, the most learning I got during this project is how as a team we need to work together, harmonize our differences, learn from each other and meet our goal. I would say this quite a journey as all of us has different backgrounds (unlike what I experienced in work whereby all of us mostly having similar experience) and we do have our work commitments but after all we all committed on our learning journey and complete this project.

Future

In term of work, I came from industry where AI not yet been the frontier of application development. So I hope to use what I learned in my course of work, most the main implementation I think will be at semantic network as this would help in many area either customer service or work efficiency.

However I do have personal goal, whereby I would like to create something that useful for the common good. A health app that the team work on for this project is something that I would like to pursue further. Covid have helped us to leap maybe a decades of technology adoption in healthcare. I do believe AI will be able to help us get a better health as it can help to reduce the problem of limited healthcare providers.

9.3.4 Personal Report: Tan Xiang Feng

Personal Contribution

For this project, I am primarily working on Strategy 3 (Detect Diabetes Retinopathy) which is something new to me. As I have developed credit risk models in my job, I have volunteered to work on the report, hoping to provide some insights on the development of model. Lastly, I have worked on the promotional video given the similarity with the report.

Lessons Learnt

I have learnt a lot from this project. We have started off the project without any market research. As such, we have encountered problem articulating the problem statement and value preposition. Subsequently, we have performed an in-depth market research, bounced off ideas with our peers and pivoted our project to better serve the needs of our target segment. From this, I have learnt the importance of market research/ fact finding and not jump directly into working on the project.

Through this project, I have my first “taste” of the deep learning framework – Tensorflow. I have built my first convolutional neural network model. While the accuracy of the model is not excellent, I am quite pleased. I have learnt about the convolutional layer, optimizer and loss function. This have better prep me for other deep learning algorithm out there.

Through this project, I have also learnt a lot from my group mates. I have no prior experience on UI/ UX and how the phone app is able to send/ retrieve information with the backend system. The project team is patient enough to help me better understand the flow and connectivity between the app and backend system. While I am still not able to build an app from scratch, I have a much better understanding on how things work.

In summary, this project has taught me the importance of market research, how to build a deep learning model and piecing of the various parts to obtaining the finished product – phone app.

Future

Through this project, I have a better understanding of system design and system implementation. I am glad to see that our phone is able to tackle the problem statement. I am also inspired to explore deeper in the field of AI and hopefully my future works are able to solve real world problems.

9.3.5 Personal Report: Zhang Zhewei

Personal Contribution

Model development and testing

Our project has 3 models based on different strategies to provide different kinds of services to customer. I'm more focused on the 1st strategy which is based on people's diet and lifestyle data make a prediction on the possibility of getting diabetes. To finish this part, I did below several steps.

- **Data processing**
  Our raw data was collected from US CDC BRFSS system, they are two over 1GB CVS files with the data collected from 2013 and 2015, and each of the file have 330 columns. After removed the incomplete and duplicate data, I picked 22 diabetes related features. After that I created two balanced datasets for model training. One is for binary classification; another one is for multi-classification.
- **Feature correlation research**
  With the feature and diabetes correlation comparing, I dropped another 5 features which are not directly impact diabetes.
- **Train/Test model**
  Because some of the feature has wide value range, so before the model training, scaling is applied on the dataset. Then I compared the performance among KNN, Naïve Bayes, Decision Tree, and Logistic Regression. Finally, I chose Naïve Bayes because the score of training and testing are close and there is less overfitting.

Project report writing

Wrote the report for our 1st model based on the above 3 steps.

Create animation video

With the purpose to engage customer and intrigue their interests, I created an animation video which is the first part of our marketing video.

Lessons Learnt

As a product manager, I participated in many projects and witnessed them launched. But this is the 1st time I participated into the implementation of an intelligent system in person. There is a lot learned during the whole progress of this project. Here I just pick some more impressive points for me.

- **Model selecting is important, but data processing is even more important.**
  At beginning I tried to use the dataset with 3 diabetes classes (no-diabetes, pre-diabetes, diabetes) to train the model, and the test score of the model is 0.82. Looks quite good, but when I put the pre-diabetes data into the model, the result is always not accurate. After some research, I found the reason is the pre-diabetes only take a small portion (1.82%) of the whole data - the dataset is quite unbalance. The unbalance data will make the result biased, especially for the

model based on similarity such as KNN. Finally, I created balanced dataset to train the model, the actual testing result is much better.

- **Take some time to research the data will have better understand and it's helpful to improve the model.**

  After data cleaning, I tried to use the dataset train the model. But the result score is not ideal, it's between 0.5 to 0.6. One of the reason is the data reliability, I found the feature of Cholesterol Check is only when people checked in 5 years will have the value 'Yes'. But some of the people didn't check in 5 years, their HighChol value is still 'Yes', these data look not reliable. So, I removed the data without check in 5 years, but HighChol is 'Yes'. Then the result do have some improvement.

Meanwhile, during the modelling selecting, I have better understand of the pros and cons of each model. Except those data processing and AI related knowledge, I also learned some software architect and UI knowledges from my teammates. It's really a good experience with a fruitful learning outcome.

Future

With the mature of AI technology, more companies include AI in their digital products. My company is also looking for better solution to improve our banking products with AI technology. Currently we are solutioning to implement smart search and smart quick links with AI technology in our next generation Internet Banking. Meanwhile, I'm also thinking using statistic model to do channels classification on our IB and P-web 5-star user feedback instead of keywords matching. In future, when more and deeper knowledge learned from the rest semesters, believe I will have a better view to use AI to improve our digital products.