



Faculty Of Computer Science and Information Technology

*A Mobile App to Predict Disease Based on Symptoms Using Artificial  
Intelligence*

H'ng Sheng Wei

Bachelor's Degree of Multimedia Computing

2022/2023

**TITLE**

**A Mobile App to Predict Disease Based on Symptoms Using Artificial Intelligence**

**NAME**

**H'ng Sheng Wei**

**69803**

This project is submitted in partial fulfillment of the requirements for the degree of  
Bachelor of Computer Science and Information Technology

Faculty Of Computer Science and Information Technology

UNIVERSITY MALAYSIA SARAWAK

2022/2023

## **TAJUK**

**Ciptaan Aplikasi untuk Meramal Penyakit Berdasarkan Simptom Menggunakan  
Kepintaran Buatan**

## **NAMA**

**H'ng Sheng Wei**

**69803**

Projek ini merupakan salah satu keperluan untuk Ijazah Sarjana Muda Sains Komputer  
dan Teknologi Maklumat

Fakulti Sains Komputer dan Teknologi Maklumat

UNIVERSITI MALAYSIA SARAWAK

2022/2023

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

**TITLE** *A MOBILE APP TO PREDICT DISEASE BASED ON SYMPTOMS USING ARTIFICIAL INTELLIGENCE*

**ACADEMIC SESSION:** YEAR 2022/2023

(CAPITAL LETTERS)

hereby agree that this Thesis\* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [ or for the purpose of interlibrary loan between HLI ]
5. \*\* Please tick ( √ )

- CONFIDENTIAL** (Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)
- RESTRICTED** (Contains restricted information as dictated by the body or organization where the research was conducted)
- UNRESTRICTED**



(H'NG SHENG WEI)

Validated by

(DR. CHAI SOO SEE)

Permanent Address

**AA3-2,  
TAMAN TUN SARDON 11700,  
GELUGOR, PENANG**

Date: 21.7.2023

Date: \_\_\_\_\_

Note \* Thesis refers to PhD, Master, and Bachelor Degree

\*\* For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

## **Acknowledgment**

The successful and final outcome of this project or study required a lot of help and guidance from many people and I would like to thank you for the completion of my work.

First and foremost, I would like to express my gratitude to my supervisor, Dr. Chai Soo See for her helpful guidance during the entire period of the project as an inspiring, interactive, and dedicated supervisor. Her professional support and advice played a very crucial role in helping me to complete my project. Secondly, I would like to give my thanks to Dr. Wang Yin Chai for giving me much useful information regarding the final-year project and the given information has helped me to complete my project in a timely and professional manner. Next, I would like to give my thanks to my lecturers for providing relevant knowledge in the regarding project.

Next appreciation goes to my university, University Malaysia Sarawak (UNIMAS) for providing me with a study platform to gain knowledge and stand a chance to learn a lot of things. The following appreciation goes to my faculty, the Faculty of Computer Science and Information Technology, for allowing me to experience challenges in my Final Year Project, which prepared me to work in the future.

A special thanks to my coursemates and friends around who helped, guided, and expressed ideas to help me with my Final Year Project. Last but not least, I wholeheartedly give my gratitude to my family for supporting me endlessly, be it financially or emotionally.

## Table Content

	<b>Page</b>
<b>Abstract</b>	1-2
<b>1.0 Chapter 1 Introduction</b>	
1.1 Introduction/Background	3-4
1.2 Problem Statement/Research Problem	5
1.3 Scope	6
1.4 Objectives	6
1.5 Brief Methodology	7-10
1.6 Significance Of Project	11
1.7 Project Schedule	12
1.8 Expected Outcome	13
<b>2.0 Chapter 2 Literature Review</b>	
2.1 Introduction	14
2.2 Literature Review	14-29
2.3 Proposed System	29-35
2.4 Summary	36
<b>3.0 Chapter 3 Requirement Analysis And Design</b>	
3.1 Introduction	37

3.2	Functional Requirement	38
3.3	Non-Functional Requirement	38-40
3.4	Hardware Requirement	40
3.5	Software Requirement	40-41
3.6	Design Diagram	42-54
3.7	Summary	55
4.0	<b>Chapter 4 Implementation And Testing</b>	
4.1	Introduction	56
4.2	Implementation Process	56-84
4.3	Summary	85
5.0	<b>Chapter 5 Discussion And Result</b>	
5.1	Introduction	86
5.2	Dataset Source	86-87
5.3	Machine Learning Data Preprocess and Train Model	87-90
5.4	Chatbot Data Preprocess and Train Model	91-93
5.5	Result	93-97
5.6	Summary	98
6.0	<b>Chapter 6 Conclusion</b>	
6.1	Introduction	99

6.2	Limitations And Challenges	99-100
6.3	Future Work	100-103
6.4	Summary	104
7.0	<b>Reference</b>	105-106
8.0	<b>Appendices</b>	
8.1	User Manual	107-117
8.2	Feedback Form	118-120



## List Of Figure

		<b>Page</b>
Figure 1.1	The Type of Methodology (Prototype Methodology)	7
Figure 1.2	Business Canvas Model for A Mobile Application to Predict Disease Based on Symptom Using Artificial Intelligence	8
Figure 1.3	Project Schedule To Develop a Mobile Application for Disease Prediction Based On Symptom Using Artificial Intelligence	12
Figure 2.1	Existing Flowchart of Disease Prediction System	26
Figure 2.2	Existing Working Model of Chatbot on Disease Prediction	26
Figure 3.1	System Architecture on Disease Prediction Using Machine Learning Model	42
Figure 3.2	System Architecture on Chatbot Model Using Natural Language Processing	43
Figure 3.3	Use Case Diagram on A Mobile Application to Predict Disease Based on Symptom Using Artificial Intelligence	44
Figure 3.4	Class Diagram on A Mobile Application to Predict Disease Based On Symptom Using Artificial Intelligence	49
Figure 3.5	Sequence Diagram for User Registration and Login Process	50
Figure 3.6	Sequence Diagram for Disease Prediction Process	51
Figure 3.7	Sequence Diagram for Requesting Disease Details Process	51
Figure 3.8	Sequence Diagram for History Viewing Process	52

Figure 3.9	Sequence Diagram for Profile Edit Or Update Process	52
Figure 3.10	Sequence Diagram for Requesting Customer Service Contact Process	53
Figure 3.11	Wireframe On a Mobile Application to Predict Disease Based On Symptom Using Artificial Intelligence	54
Figure 4.1	Library Import In Main Python File	56
Figure 4.2	Library Import In Machine Python File	57
Figure 4.3	Library Import In Training Python File	57
Figure 4.4	Library Import In Model Python File	57
Figure 4.5	Library Import In Machine Learning Model FastAPI Python File	57
Figure 4.6	Library Import In ChatBot Model FastAPI Python File	58
Figure 4.7	Dataset (Disease With Symptoms)	60
Figure 4.8	Dataset Symptom Description (Description Of Diseases)	60
Figure 4.9	Dataset Symptom Precaution (Precaution Of Diseases)	61
Figure 4.10	Dataset Symptom Severity (Severity Of Symptoms)	61
Figure 4.11	Chatbot Model Trained (Part One)	63
Figure 4.12	JSON File Used In Chatbot Model	63
Figure 4.13	Chatbot Model Trained (Part Two)	64
Figure 4.14	Machine Learning Model Trained Using Jupyter Notebook (Data Preprocessing)	65

Figure 4.15	Support Vector Machine Model Trained Using Jupyter Notebook	66
Figure 4.16	Support Vector Machine Model Trained Result	66
Figure 4.17	Decision Tree Model Trained Using Jupyter Notebook	67
Figure 4.18	Decision Tree Model Trained Result	67
Figure 4.19	Random Forest Model Trained Using Jupyter Notebook	68
Figure 4.20	Random Forest Model Trained Result	68
Figure 4.21	K-Nearest Neighbors Model Trained Using Jupyter Notebook	69
Figure 4.22	K-Nearest Neighbors Model Trained Result	69
Figure 4.23	Gaussian Naïve Bayes Model Trained Using Jupyter Notebook	69
Figure 4.24	Gaussian Naïve Bayes Model Trained Result	70
Figure 4.25	Random Forest Model	71
Figure 4.26	Loaded Tensorflow Lite Model	72
Figure 4.27	Kivy File Of Main Page (Logging Process)	73
Figure 4.28	Kivy File Of Register Page (Register Process)	74
Figure 4.29	Kivy File Of History Page (History Process)	74
Figure 4.30	Render Dashboard	75
Figure 4.31	Render For Chatbot Model	75
Figure 4.32	Render For Machine Learning Model	76
Figure 4.33	Github For Chatbot Model	76

Figure 4.34	Github For Machine Learning Model	77
Figure 4.35	Buildozer with Google Colaboratory (Install Required File)	78
Figure 4.36	Buildozer with Google Colaboratory (Buildozer Spec File & Build APK file)	78
Figure 4.37	Buildozer Spec File In Google Colaboratory	79
Figure 4.38	Splash Screen	80
Figure 4.39	Login Screen	80
Figure 4.40	Registration Screen	81
Figure 4.41	Home Screen	81
Figure 4.42	History Screen	82
Figure 4.43	Profile Screen	82
Figure 4.44	Setting Screen	83
Figure 4.45	Change Password Screen	83
Figure 4.46	Feedback Google Form	84
Figure 5.1	Cleaning Data	87
Figure 5.2	Replace Null Or N/A Data	88
Figure 5.3	Encoding Symptom With Severity Weight (Part 1)	88
Figure 5.4	Encoding Symptom With Severity Weight (Part 2)	89
Figure 5.5	Machine Learning Model Trained	89

Figure 5.6	Initialization of Chatbot Model	91
Figure 5.7	Preprocessing The Data From Json File	91
Figure 5.8	Building Training Dataset	92
Figure 5.9	Compile Chatbot Model and Train Chatbot Model	92
Figure 5.10	Confusion Matrix Random Forest Model	95
Figure 5.11	Confusion Matrix Gaussian Naïve Bayes Model	96
Figure 6.1	Cover User Manual	107
Figure 6.2	Table Content User Manual	108
Figure 6.3	Introduction, System Requirements and Installation Instructions in User Manual	109
Figure 6.4	User Interface Overview User Manual (Part 1)	110
Figure 6.5	User Interface Overview User Manual (Part 2)	110
Figure 6.6	Using Disease 02 User Manual (Part 1)	111
Figure 6.7	Using Disease 02 User Manual (Part 2)	112
Figure 6.8	Using Disease 02 User Manual (Part 3)	113
Figure 6.9	Using Disease 02 User Manual (Part 4)	114
Figure 6.10	Using Disease 02 User Manual (Part 5)	115
Figure 6.11	Frequently Asked Questions User Manual	116
Figure 6.12	Contacting Support User Manual	117

Figure 7.1	Feedback Form (Part 1)	118
Figure 7.2	Feedback Form (Part 2)	119
Figure 7.3	Feedback Form (Part 3)	120

## List Of Table

		<b>Page</b>
Table 2.1	A Comparative Study Using Various of Medical Chatbot Techniques on Disease Prediction In Literature Review	16-17
Table 2.2	A Comparative Study Using Various of Machine Learning Algorithms on Disease Prediction in Literature Review	22-26
Table 2.3	Comparison First Three Existed Similar Mobile Applications with Upcoming Prototype	27-29
Table 2.4	Confusion Matrix	34
Table 3.1	Use Case Description for Chat Conversation	45
Table 3.2	Use Case Description for Registration And Login	45-46
Table 3.3	Use Case Description for Input Symptom	46-47
Table 3.4	Use Case Description for View Prediction Result	47
Table 3.5	Use Case Description for View Prediction History	48
Table 3.6	Use Case Description for View Profile Details	48
Table 4.1	Dataset Source	62
Table 5.1	K-Fold Cross Validation Score of Machine Learning Model	93
Table 5.2	Accuracy, F1, Precision and Recall Score of Machine Learning Model	94

## **Abstract**

As we knew nowadays, the world was surrounded by different types of diseases, and they caused humans to live in fear of disease and even brought death, especially of Coronavirus Disease (Covid-19). Therefore, it is important to develop a disease prediction to give early detection of diseases that might be infected by humans. The healthcare department has been doing much research in the fields of intelligent consultation, disease diagnosis, intelligent question-answering doctors like AI chatbot and so on. This had made many achievements. To improve medical technology, this study intends to use healthcare data analysis combined with machine learning knowledge to provide patients with a simple disease prediction which is usually neglected for lacking professional knowledge of the disease. This helps patients to get a suitable way of treatment in a short time before their health condition gets worse and worse. A different suitable machine learning algorithm will be used in the prediction system to predict the disease based on the symptoms of patients. To reduce time, the Chabot feature was used too so that patients were able to save time without meeting for a doctor to get treatment. The result at the end will show that our approach improved the accuracy of disease prediction based on symptoms with different evaluation metrics.



## **Abstrak**

Seperti yang kita ketahui sekarang, dunia dikelilingi oleh pelbagai jenis penyakit, dan ia menyebabkan manusia hidup dalam ketakutan terhadap penyakit malah membawa kematian terutamanya Penyakit Koronavirus (Covid-19). Oleh itu, hal ini adalah penting untuk membangunkan ramalan penyakit untuk memberikan pengesanan awal penyakit yang mungkin dijangkiti oleh manusia. Jabatan penjagaan kesihatan telah banyak melakukan penyelidikan dalam bidang perundingan pintar, diagnosis penyakit, doktor menjawab soalan pintar seperti AI chatbot dan sebagainya. Ini telah membuat banyak pencapaian. Untuk meningkatkan teknologi perubatan, kajian ini berhasrat untuk menggunakan analisis data penjagaan kesihatan digabungkan dengan pengetahuan pembelajaran mesin untuk memberikan pesakit ramalan penyakit yang mudah yang biasanya diabaikan kerana kekurangan pengetahuan profesional tentang penyakit. Ini membantu pesakit mendapatkan cara rawatan yang sesuai dalam masa yang singkat sebelum keadaan kesihatan semakin teruk. Algoritma pembelajaran mesin yang berbeza akan digunakan dalam sistem ramalan untuk meramal penyakit berdasarkan simptom pesakit. Untuk mengurangkan masa, ciri Chabot juga digunakan dan supaya pesakit dapat menjimatkan masa tanpa berjumpa doktor untuk mendapatkan rawatan. Keputusan pada akhirnya akan menunjukkan bahawa pendekatan kami meningkatkan ketepatan ramalan penyakit berdasarkan gejala dengan metrik penilaian yang berbeza.

## **1.0 Chapter 1 Introduction**

### **1.1 Introduction/Background**

In the new generation, the world was surrounded by different diseases that might cause difficult issues for humans in their daily life and this pandemic also caused many problems occurred in the fact of economy. One of the pandemics, Coronavirus Disease (Covid-19) had spread throughout the world since the year 2019 and it caused many death cases along with the statistics that had been analyzed by a specific department. The best way to prevent and slow down transmission is to be well-informed about the disease and how the virus spreads. Learn to protect yourself and others from infection by staying at least 1 meter apart from others, wearing a properly fitted mask, and washing your hands or using an alcohol-based rub frequently. We should get vaccinated when it was our turn to do so and follow local guidance. Even now although Covid-19 was secured by vaccine injection, the disease was not cured and might infect again by showing some the symptoms like fever, cough and so on. This might bring humans difficulty in their daily life by always wearing the mask outside the home and not being sure the of exact symptoms. Therefore, technology to predict the disease might need for humans to easily get know what disease once they got infected anywhere and immediately request treatment.

Artificial intelligence (AI) knowledge may bring a large change and benefits to humans, especially for the medical department to solve the issues of the spread of disease and pandemics before becoming worse and worse. It refers to the simulation of human intelligence in machines that are programmed to think like humans such as “learning” and “problem-solving” and mimic their actions according to their reflection (Divya S. et al., 2018). For AI, the chatbot will be used in the prototype to automatically do predictions on the disease and reply to the users after inputting the symptoms. The technology used to build a chatbot mostly was using Natural Language Processing (NLP) or Natural Language Understanding (NLU) to verify and understand each input. Chatbots can work as a virtual doctor so that patients will not always face the difficulty of making an appointment with a doctor physically. This helps in saving more time in getting the treatment immediately and might help doctors in decreasing their workload. Machines are always considered

better than humans as they might be without any human error, they can perform the tasks more efficiently and with a consistent level of accuracy (Keniya R. et al., 2020). This may help a lot to us in our daily life. Most of the models were initiated by using various machine learning (ML) algorithms that collected raw data and then categorized it according to gender, age group, and symptoms. This model was used to undergo the process of disease prediction and the ML algorithm was important to increase the accuracy of prediction. A dataset will be also needed to get the data o information on different symptoms and it can be obtained from the Kaggle website as it was free of charge to continue the process of disease prediction. After that, the information collected might be used to develop a mobile application to predict the disease based on specific symptoms throughout the process of the project and it can be used anywhere.

## **1.2 Problem Statement/Research Problem**

Nowadays, in the new generation people are living in a world that was plagued by pandemics such as Covid-19. In many cases, people sometimes will have difficulty requesting an appointment to see the doctors and get doctors' help. This caused many people cannot to identify their disease and take medicines without consulting doctors. Therefore, technology to predict disease based on symptoms was developed to give the potential users peace of mind that they can get treatment immediately.

Besides that, healthcare industries have a massive number of databases consisting of different types of data such as structured, semi-structured or unstructured. This was a challenge to them for managing the huge dataset and different data types or attributes without any errors occurring. The application of predictive analytics in the healthcare sector has received a great amount of interest in the research community (Krishnamoorthi R. et al., 2022). Machine learning algorithms are now mostly used in current modern technology especially in fact to increase the accuracy of prediction. This might help to organize the huge dataset easily and efficiently using machine learning algorithms.

### **1.3 Scope**

In this project, the scope was to let humans able to undergo a simple disease prediction based on their symptoms so that they can get simple treatment if it was an emergency issue and avoid it becoming worst, even if the disease may cause death. Besides that, the world now a day was surrounded by different diseases that might cause dangerous issues for us. Therefore, technology to help in disease prediction should be developed and improve the current trend in the medical department.

### **1.4 Objectives**

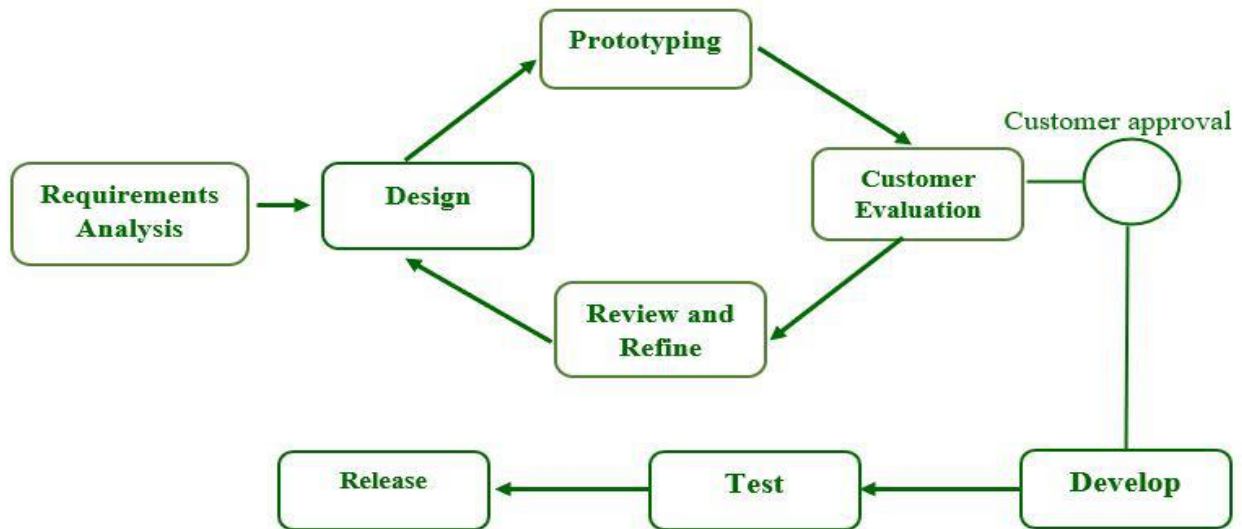
**1.4.1** To design a system for a prediction of the disease based on the symptoms by using artificial intelligence including some machine learning algorithms.

**1.4.2** To compare the differences between machine learning algorithms and select the best machine learning algorithms for disease prediction.

**1.4.3** To develop a mobile application that can be able to access anywhere for undergoing the disease prediction process.

## 1.5 Brief Methodology

The type of methodology that will be used in this project was Prototype Methodology.



**Figure 1.1:** The Type of Methodology (Prototype Methodology)

Prototype methodology consists of different phases starting with requirement analysis, design, prototype, users' evaluation and last comes with the review and update phase. Once get the approval from the users' evaluation phase, then it will come for the development, test and maintain or release phase process.

### 1.5.1 Requirement Analysis

In the first phase which was the requirement phase, some analysis like comparing some existing mobile applications and problem analysis will be done through some method before continuing to another phase. The method to complete the analysis and full fill our objectives was using the Business Model Canvas (BMC) might be applied. Besides that, the functional, non-functional, hardware and software requirements have to be analyzed in this phase too before continuing to another phase. The model that consists of a machine

learning algorithm will be analyzed to see which algorithm was suitable for the project to increase the accuracy of prediction. A few examples of machine learning algorithms such as K-nearest neighbors (KNN), Naïve Bayes, Gaussian Naïve Bayes, Decision Tress and so on.



**Figure 1.2:** Business Canvas Model for A Mobile Application to Predict Disease Based on Symptom Using Artificial Intelligence

### 1.5.2 Design

To start the phase of design, the wireframe of the prototype uses a specific platform or software such as Figma to give us a clear on what feature and how the design should contain in our mobile application that to predict disease based on symptom. Some of the features

such as a conversation between AI chatbot, disease prediction based on user's symptoms, registration and login for the new users, verification of users and so on should contain in the prototype and needed to draft out as guidance for our real output on the prototype.

### **1.5.3 Prototype**

In this phase, the prototype will be developed by using mostly Python programming language and the software might be Phycharm or other suitable software and platform such as Jupyter Notebook. Besides that, the knowledge of Artificial Intelligence will be applied to develop an AI chatbot for automatic response to disease prediction. Then it might use the Linux Oracle box method that installed Ubuntu to generate the output as a mobile application that can be available on our mobile phone and use it. The machine learning model will be trained and tested using a machine learning algorithm and the purpose to do so is to increase the accuracy of prediction on disease based on the user's symptoms. Before this, the dataset will be undergoing data pre-processing to avoid missing values, duplicated data or other data issues.

### **1.5.4 Evaluation**

For this phase, a form will be provided to get some suggestions and feedback from upcoming users to see which features can be considered or improved in the prototype. The form should be analyzed to see the probability of the human need for this kind of technology in their life. Then, the coming process which were development, test and maintain or release will be needed to undergo after the process of evaluation of users.

#### **1.5.4.1 Development**

In the development phase, once get the approval from the evaluation of users then the prototype will be developed to output as a mobile application. But in this phase, it was still testing on the mobile application in purpose to get more suggestions and feedback from the users.