# WAITING TIME PREDICTION SYSTEM USING LINEAR REGRESSION METHOD

# SYAMSUL IKHMAL BIN IBRAHIM



# UNIVERSITI TEKNIKAL MALAYSIA MELAKA

### WAITING TIME PREDICTION SYSTEM USING LINEAR **REGRESSION METHOD**

### SYAMSUL IKHMAL BIN IBRAHIM

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Electronic Engineering with Honours

2022

**Faculty of Electronics and Computer Engineering** Universiti Teknikal Malaysia Melaka

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

a.

# DECLARATION



# APPROVAL

Date : 21 JUNE 2022

### DEDICATION

A special thanks to all those who have supported, encouraged and inspired me, especially to my honourable supervisor, beloved family, and friends for all their guidance, love and attention which has made it possible for me to make it up to this point. I also like to thank the final year project (FYP) committees who organised the talks to share and guide the students in developing the project and FYP project.

اونيوم سيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

### ABSTRACT

Malaysia now has a wide range of dental facilities. In reality, the number of patients does not decrease because treatment in government clinics is still far less cost than treatment in private clinics. This project is about developing a management queue system that will solve current problems that dental clinics are experiencing, especially government clinics. In this study, prediction systems using 2 machine learning techniques, Linear Regression and Random Forest were developed and analysed to overcome this difficult scenario. This model was developed in Python using Jupyter Notebook. To examine the performance of this machine learning technique, the regression metrics Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used. A previously selected dataset from dental clinics was used to predict the duration of queueing in real life for this method. As a result, the RMSE and MAE values were shown in a table result, and Linear Regression has lower RMSE and MAE values than Random Forest, indicating a good machine learning performance model. The dataset's waiting time is displayed in minutes by developing a graphical user interface (GUI).

### ABSTRAK

Malaysia kini mempunyai pelbagai kemudahan pergigian. Realitinya, bilangan pesakit tidak berkurangan kerana rawatan di klinik kerajaan masih jauh lebih murah berbanding rawatan di klinik swasta. Projek ini adalah untuk membangunkan sistem barisan pengurusan yang akan menyelesaikan masalah semasa yang dialami oleh klinik pergigian, terutamanya klinik kerajaan. Dalam kajian ini, sistem ramalan menggunakan 2 teknik pembelajaran mesin, Regresi Linear dan Hutan Rawak telah dibangunkan dan dianalisis untuk mengatasi senario sukar ini. Model ini dibangunkan dalam Python menggunakan Jupyter Notebook. Untuk mengkaji prestasi teknik pembelajaran mesin ini, metrik regresi Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE) telah digunakan. Set data yang dipilih sebelum ini daripada klinik pergigian telah digunakan untuk meramalkan tempoh beratur dalam kehidupan sebenar untuk kaedah ini. Hasilnya, nilai RMSE dan MAE telah ditunjukkan dalam jadual hasil, dan Regresi Linear mempunyai nilai RMSE dan MAE yang lebih rendah daripada Hutan Rawak, menunjukkan model prestasi pembelajaran mesin yang baik. Masa menunggu set data telah dipaparkan dalam minit dengan mencipta antara muka pengguna grafik (GUI).

### ACKNOWLEDGEMENTS

First and foremost, I would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for giving me this opportunity to practice the knowledge I had gained throughout the years of study in a real-life situation.

Next, I also like to thank the final year project (FYP) committees who organised the talks to share and guide the students in developing the FYP project. Special gratitude to my final year project supervisor and co-supervisor, Dr Noor Asyikin Binti Sulaiman and Dr Noorazlan Shah bin Zainudin, who helped coordinate my project and gave me enough suggestions and encouragement. The project developed successfully because of their guidance and constant supervision. I appreciate the advice from the panels in my project presentation, which has enhanced my presentation skills, and I am thankful for their comments and recommendations.

Besides, I would like to express my gratitude to my parents, who always gave me support and encouragement in many ways. I also want to thank my siblings, who cooperate with me and are generous in sharing their knowledge about this project. Lastly, I would also like to thank my friend and colleagues who have given me help during the development of my project and the report.

# **TABLE OF CONTENTS**



1.3	Objectives	4
1.4	Scope of Project	4
1.5	Project Significant	5
CHA	APTER 2 BACKGROUND STUDY	6
2.1	Queueing System	6
2.2	Machine Learning	10
2.3	Machine Learning in Waiting Time System	11
	2.3.1 Linear Regression	15
	2.3.1.1 Visualising and a fitting a Linear Regression model	18
	2.3.2 Random Forest	19
	2.3.3 Support Vector	20
	اونيومرسيني تيڪنيڪAveural Network 2.3.4	21
2.4	Linear regression in the waiting time system SIA MELAKA	22
CHA	APTER 3 METHODOLOGY	24
3.1	Overview of the project	24
3.2	Project Implementation	27
3.3	Data collection	27
3.4	Data cleaning	28
3.5	Feature selection	28
3.6	Split data	30

v

3.7	Train Linear Regression model	30	
	3.7.1 Making Predictions from Linear Regression models	31	
	3.7.2 Root Mean Square Error	38	
	3.7.3 Mean Absolute Error	38	
3.8	Comparison of Machine Learning Models	38	
3.9	Development of Graphical User Interface (GUI) for Oueue System for		
	Waiting Time Prediction	39	
CHA	APTER 4 RESULTS AND DISCUSSION	41	
4.1	Linear Regression Model	41	
	4.1.1 Performance of Linear Regression Model	43	
	4.1.2 Comparison with Random Forest Regression Model	44	
4.2 Queue System for Waiting Time Prediction using Developed Graphics			
	اونيونر سيتي تيڪنيڪل ملي(GUI)	44	
	4.2.1 Testing of an interface of Queue System for Waiting Time		
	Prediction	44	
4.3	Environmental and Sustainability	46	
4.4	Chapter summary	46	
CHAPTER 5 CONCLUSION AND FUTURE WORKS			
5.1	Conclusion	48	
5.2	Future Works	49	
REF	ERENCES	51	

vi

# LIST OF FIGURES

Figure 2.1: Hospital Queue Management System [5]	8
Figure 2.2: Comparison of R2 prediction among 4 modalities [17]	13
Figure 2.3: Explanation of linear regression	16
Figure 2.4: Types of relationship	17
Figure 2.5: Type of relationship continued	17
Figure 2.6: Type of relationship continued	18
Figure 2.7: Relationship between the observed value and fitted value	19
UNIVERSITI TEKNIKAL MALAYSIA MELAKA Figure 2.8: Random Forest Model	20
Figure 2.9: Two classes of data point separated by a hyperplane	21
Figure 2.10: Neural Network	22
Figure 3.1: Flowchart of the project	26
Figure 3.2: Patient records raw dataset	27
Figure 3.3: Split Dataset	30
Figure 3.4: Import Python library	32
Figure 3.5: Clinic dataset describes	32
Figure 3.6: Replace characters with numerical values	33

Figure 3.7: Clinic dataset	33
Figure 3.8: Build the linear regression model	34
Figure 3.9: Evaluation regression metric values	35
Figure 3.10: Predicted value	35
Figure 3.11: Theoretical calculation value	36
Figure 3.12: Independent variables	36
Figure 3.13: Dependent variable	37
Figure 3.14: Build a Linear Regression model	37
Figure 3.15: GUI of Queue System for Waiting Time Prediction	40
Figure 4.1: Linear Regression Model	42
Figure 4.2: Testing of GUI for prediction	45
اونيۆم سيتي تيڪنيڪل مليسيا ملاك	
UNIVERSITI TEKNIKAL MALAYSIA MELAKA	

viii

# LIST OF TABLES

Table 3.1: List of data types	29
Table 3.2: Python library and function	31
Table 4.1: Linear Regression of regression metrics	43
Table 4.2: Random Forest of regression metrics	44
اوييوم سيتي بيڪنيڪل مليسيا ملاك	
UNIVERSITI TEKNIKAL MALAYSIA MELAKA	

### LIST OF SYMBOLS AND ABBREVIATIONS



### **CHAPTER 1**

### **INTRODUCTION**



This chapter will discuss more in detail the project in the general background, problem statement, objectives, the scope of the project and project significant.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

#### 1.1 Project Background

Over the years, waiting in line has become one of the most inconvenient tasks all over the world. Queues are something that almost everyone experiences regularly. Queues are common in our daily lives and are a part of services, such as waiting for treatment in a clinic, waiting for groceries, waiting for cash out at an automated teller machine (ATM), and so on. Queuing issues also arise when multiple customers require a source of information and the service is unable to fulfil the level of demand. Nowadays, technology is rapidly advancing in all fields, including healthcare. The clinic's system's development will lead to better medical equipment and reduced complexity.

Following that, Malaysia has a variety of dental clinics. People who are sick will go to dental clinics, especially government clinics. The main reason why some people prefer to seek treatment at government clinics is that the cost of treatment is lower than in private clinics. However, the number of patients is not reducing even during the Covid-19 pandemic. While waiting for treatment, a patient has a lack of information about the waiting time to get treatment and unexpected delays that can be highly frustrating in a stressful environment.

Practically, government dental clinics in Melaka are still using the traditional queue system, which is the take number system. Typically, the patient must wait up, register, and then receive a ticket queue. The time difference between treatments would not be consistent since the medical management would differ. The time is determined by the doctor, which estimates how much time is required to complete each treatment, such as consultation, treatment, follow-up appointments, and so on. Nowadays, in waiting time prediction, artificial intelligence is used to estimate the patient's waiting time by employing a variety of machine learning techniques and comparing which model technique is particularly well suited.

The purpose of this project is to predict patient's waiting times at the dental clinics. Hence, the development of Machine Learning in clinic management gives an effective and well-organised queueing process. Machine Learning is a sort of artificial intelligence (AI) that enables software applications to become more accurate at predicting outcomes without explicitly programming them to do so. Linear Regression is a machine learning approach that uses a straight line to estimate

the relationship between the independent variable (input) and the dependent variable (predict). As previously stated, the data from Klinik Pergigian Alor Gajah in Melaka was used to develop this system. This study helps in time estimation and attempts to reduce waiting times in dental clinics, hence reducing crowds. Patients are given an estimated time for treatment so that they can effectively manage their time while waiting.

#### **1.2 Problem Statement**

Government clinics are crowded places where people will come every day since the cost of treatment in government clinics is much cheaper than in private clinics. They must deal with the long delay at the same time. If we are physically and emotionally stable, the waiting period would be normal. When it comes to the patient, things go in the opposite direction. Patients are mentally and physically unstable, as well as weak. Usually, the waiting area is crowded, and patients are asked to stand nearby or sit in a different place and come after a few minutes to check on their turn. It is difficult for the patient to move anywhere, and there is no way to estimate the waiting time for new cases, appointments, follow-ups, and other treatments.

Nowadays, our government is implementing a social distancing environment, where the waiting area is no longer enough space for all the patients. The covid-19 pandemic has made the situation worsen to the government hospitals are facing an unexpected number of patients. A web application is one of the latest technologies that can browse or read the pages, fill out forms, register for participating in transactions, and download and save the pages. Unfortunately, it must have a good internet connection to access it and also it will use the money to purchase the data connection to access the internet. If the website has some problems, it will also affect the website in web applications. It is pretty troublesome for the patients.

#### 1.3 Objectives

The main objectives of this project are :

- i. To develop a time prediction system using the linear regression method.
- ii. To design a Graphical User Interface (GUI) that is much easier for the user to interact with it.

#### 1.4 Scope of Project

This project aims to develop a queue system that will assist to predict waiting times in dental clinics. This project comprises Python Programming with the machine learning library for Python language. It consists of various algorithms and supports scientific libraries. Data pre-processing is the procedure for preparing suitable raw data in a machine learning model. Then, the dataset is normalised so that the dataset's variables lie within a specific range before splitting into training, and testing datasets. Linear Regression, one type of machine learning model algorithm is applied in this work. The performance of the prediction models is evaluated by computing the regression metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The important parameters are also identified and proposed with the best prediction model. Lastly, the development of GUI for the system is also be done for easy access to the dental clinic.

#### **1.5 Project Significant**

The objective of this project is to develop a time prediction system using the linear regression method. This is because linear regression is a much simpler technique than other machine learning techniques. It improves the estimation technique and, more crucially, these techniques have simple equations and an easy-to-understand interpretation on a modular level. Second, the objective of this project is to design a graphical user interface (GUI). This project makes it much easier for the user to engage with it, such as inserting data into the GUI using the instructions provided.

This project has the opportunity of commercialization because this system can be used in any clinic. This system is easier for patients to know the expected time to get their treatment from the doctor. A GUI has a simple design and is much easier to use. The user does not need to understand everything inside the GUI, such as data cleaning, data splitting, machine learning model error and accuracy, and so on.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

### **CHAPTER 2**

### **BACKGROUND STUDY**



This chapter describes the phenomenon of queuing that happened not necessarily at the dental clinic but in other healthcare too. The literature related to the queueing waiting time system that used machine learning to predict the time of waiting lines. It is then followed by the type of machine learning used to improve the waiting time system.

#### 2.1 Queueing System

Queuing is a common activity that people experience in their daily lives. Queue occurs when the demand for a service exceeds the total supply [1]. In our lives, we spend our time queuing up to meet our needs, even if it consumes much time. A queue is a waiting time that starts from when the customers get their number until

they enter the consultation room for examination [2]. This scenario was likely to occur in healthcare sectors such as hospitals and clinics with higher customer presence. The customer or patient needs to wait for their number to be called up, where the numbers will be displayed and voiced out. Regardless of their level of health conditions, they will be gathered in a specific area or waiting room together with a printed slip paper while waiting for their turn to be called up. This will cause the waiting area to be bustled and uncomfortable.

Based on Ministry of Health Malaysia policies [2], the standard of waiting time for outpatients to get a consultation from a doctor is within 30 minutes except for emergency cases. A study found that in 21 hospitals in Malaysia, the average waiting time for the outpatient department to see a doctor is one hour, and they need to wait for another hour to take their medicine [3]. While a study has analysed that the waiting time to get a consultant from a doctor more than 2 hours will lead to patient dissatisfaction, thus affecting the hospital or clinic performance [4]. Such a result was interpreted where many factors contribute to patient dissatisfaction and clinic performance, such as lack of physicians, staff, facilities, etc. Figure 2.1 shows the typical hospital queue management system where patients need to wait in the waiting area without going anywhere while waiting for their turn to be called. As a result, they cannot simply go anywhere because they may miss their turn due to a lack of information regarding the queue's progress and, at the same time, their time was simply wasted. Therefore, the waiting time in healthcare needs to be improved to provide convenient services to the patient.



Figure 2.1: Hospital Queue Management System [5]

In order to enhance the traditional physical queue at the hospital, Hospital Service Queue System has been proposed [5]. This system aimed to help the public hospital that suffers from long queues. The system managed the hospital queue online, where **UNVERSITIEKNIKAL MALAYSIA** customers, patients and stakeholders could access their queues remotely via a web application. The advantage of this system is that it allows people to monitor their queuing status and provides estimated time services through their smartphone via internet access. However, the proposed system not has been tested for real data thus, there was no experimental result reported.

Naim et.al [6], have proposed a push notification mobile application to alert the user when their number is called through the mobile device. The user needs to press the queueing device provided to get the queue number with a printed QR code during the registration at the counter. The push notification is triggered by the web

application and Firebase notification. Despite that, the system cannot be used if the user's smartphone is out of battery or does not have internet connectivity. Therefore, the user needs to keep the printed queue number with them, and the queue number must be displayed on a monitor.

Wristbands for E-queue hospitals are designed for patient queuing at clinics/hospitals, especially for mothers carrying infants, elderly people, and patients who do not have a companion in the clinic or hospitals [7]. This project aims to improve a queueing system thus making the waiting area less crowded. For this system, each patient should have an RFID tag which consists of their detail and they need to scan their tag so that the patient's information will automatically key into the hospital registration system. The wristband also will be given during registration. However, the doctor needs to register the Blynk application first on their smartphone to get an email notification of Auth Token to activate the wristband. All the databases of the patient will be stored in the system and Blynk application. The doctors will select the next patient's name and a particular wristband would be triggered and it will display the patient's name and the room number. The Wristband consists of the buzzer as a sound or mini vibrating motor as vibration which to the patient and this situation depends on the patient's capability. Unfortunately, this system must be within the Wi-Fi range, and if the patient is outside of the Wi-Fi range, the patient needs to register again. Therefore, there is still another way that might be encountered for Wristbands for the E-queue project. We can estimate the waiting time by developing a time prediction system using artificial intelligence. Due to advances in technology, this project has the potential to go further. Some projects do not require internet access and also can be completed using an offline system.

#### 2.2 Machine Learning

Machine learning (ML) is a general term for computational algorithms that use the experience to enhance performance or generate accurate predictions. Machine learning is a subset of artificial intelligence (AI) that enables a software programme to anticipate outcomes without explicitly programming them. [8]. Machine learning is used to educate machines on how to handle data effectively. Thus, this algorithm uses historical data to predict new output values. The goal of machine learning generally is to understand the structure of data and fit that data into models that people can understand and utilise. Nowadays, most technology used today has benefitted from machine learning. Machine learning is continuously used in many fields such as business, healthcare, forecasting, fraud transaction, financial management, spam detection and so on. Many previous papers have used machine learning techniques in their projects. Firstly, Satya Hermanto et.al [9], have applied Artificial Neural Network (ANN) to investigate predictive variables that explain waiting-time duration in bank customer queues. Other from that, Sulaiman et.al [10], have used three supervised learning classifiers such as Deep Learning, Support Vector Machine and Multi-Layer Perceptron in classifying in terms of six different classes of fault detection for air conditioning system. Lastly, Pang et.al [11], were successfully implemented 4 types of machine learning techniques such as Random Forest, Linear Regression, Decision Tree and Neural Network to predict rubber crop production.

#### 2.3 Machine Learning in Waiting Time System

Waiting in a long queue can be stressful and exhausting for clients. Many queueing waiting time systems have been developed to predict the clients waiting time. To predict the waiting time for services, queuing theory analysis is needed because it helps describe features of the queue and the design of the services process. Various prior research uses machine learning to predict the queue's waiting time as this technique gives high accuracy of prediction.

Athanasios et.al [12] adopted a queueing model with a neural network machine learning algorithm to estimate the waiting time of customers served in a bank. Through this research paper, a few parameters were presented to predict the estimated waiting time of the clients. A neural network was chosen over other machine learning models due to the continuous training capabilities of a neural network. Besides, Anussornnitisarn and Limlawan [13] proposed the queue prediction system using an Artificial Neural Network (ANN) to predict the waiting time. The proposed system can monitor and detect changes in queue length. This algorithm resulted in improving the accuracy of waiting time prediction for service to the customers. Satya Hermanto et.al [14] presented a customer queue wait time using the ANN algorithm to predict the queue at the bank. Based on this research, this algorithm was used as the queue at the bank is unpredicted and the customer left the queue before being served, hence affecting the queue. This approach is utilised so that the waiting time at the bank can be predicted as accurately as the actual waiting time.

Apart from that, Roberto et.al [15] proposed a predictive model which to predict waiting time overflow in bank queues. The predictive models provide a probability of Time Overflow whenever a new customer enters the queue. Thus, this model use queueing theory formula to predict waiting time on single server queues. Gradient Boosting machine learning (GBM) and several machine learning algorithms has been trained. GBM model resulted in the highest metric compared to other algorithms. To find significant results in predicting waiting time on queue, Gomes et.al [16] implemented a Support Vector Regression algorithm. Nevertheless, the algorithm resulted in not high accuracy, which has a huge gap compared to the actual waiting time.

In this project research [17], the authors studied the applicability of machine learning models to predict waiting times at a walk-in radiology facility (radiography) and delay times at scheduled radiology facilities in the Massachusetts General Hospital Department of Radiology. The author uses the extracted 9 principal examination parameters: "patient arrival time, examination begin and complete times, time of the first image acquisition, examination code, examination description, scanner name, modality, and division of examination". Several machinelearning algorithms, such as Neural Network, Random Forest, Support Vector Machine, Elastic Net, Multivariate Adaptive Regression Splines, K-Nearest Neighbors (KNN), Gradient Boosting Machine (GBM), Bagging, Classification and Regression Tree (CART), and Linear Regression, were evaluated to find the most accurate method. Based on the result, prediction accuracy in the form of R-Squared (R2) was different among modalities. In this study, the author defined that the elastic net model performed best among the 10 proposed models for predicting waiting times or delay times across all four modalities as tabulated in Figure 2.2.

Method	СТ	MRI	Ultrasound	Radiography
GBM	0.3084	0.2646	0.5895	0.4483
Elastic net	0.3220	0.2971	0.5630	0.4642
Random forest	0.2959	0.2470	0.5790	0.4582
Linear regression	0.3205	0.2980	0.5626	0.4631
MARS	0.3196	0.2541	0.5793	0.4520
SVM	0.3071	0.2694	0.5952	0.4543
Bagging	0.2384	0.2215	0.5016	0.4414
CART	0.1723	0.1318	0.2115	0.3793
Neural network	0.2003	0.1706	0.4767	0.4475
KNN	0.0742	0.1388	0.4177	0.4287

Figure 2.2: Comparison of R2 prediction among 4 modalities [17]

One of the main issues nowadays for bus operators is that the Estimate Time of Arriving (ETA) is not accurate and it deviates from the actual ETA by too much, and this discourages riders and so ridership is affected in the long run. Rafidah et.al [18] proposed a machine learning model that may provide a more accurate ETA. This research project aimed to predict bus arrival time based on the ETA of the bus at departure from one station to the next station by incorporating weather conditions. Support Vector Machine classifier model which is Support Vector Regression (SVR) was used in this research. In this study, the data used was Urban City Bus data which covers part of the Petaling Jaya area (route name PJ03). Input features for the training of the model include segments, distance, weather, and peak or non-peak hour are used to predict the travel duration for the segments. The data has been categorised into peak and non-peak hours based on the time the record is logged. Different types of hourly weather data have been assigned to each of the samples as well based on the log time. The experimental result indicates the SVR model displays good prediction ability with its low average error on the prediction result compared to the previous study. Overall, SVR may be a feasible model for ETA prediction but it needs to be trained and tested vigorously with more data and features.

Y. Sanit-In and K. R. Saikaew [19], proposed and evaluated the model for the waiting time prediction in one-stop services. Two types of data were tested: queue logs of Khon Kaen University post office and queue logs of the ear nose and throat clinic, Srinagarind Hospital. Three approaches have been implemented and compared including Queueing Theory using M/M/c pattern, Average Time prediction, and Random Forest algorithm. Based on both experimental results, Random Forest algorithm is the most precise and effective model compared to Queuing theory while Average Time has the least accuracy. As a result, Random Forest model achieves the highest accuracy because Random Forest is a supervised learning algorithm that differs from Queueing Theory and Average Time which are based on mathematical formulas. Regarding the mean result, the number of waiting queues is the most effective model because the number of waiting queues is the predictive model because the number of waiting queues is the feature that indicates the quantity of the service at the time.

However, a no-show is when a patient misses an appointment that was previously scheduled. This phenomenon happens in all sorts of areas, where there is the need to schedule patients or clients into a time slot. As mentioned in [20], this paper aims to create a prediction model that automatically returns the no-show probability for an appointment, and can efficiently test data from many clinics and hospitals without the need of an advanced user to improve the system. Findings causes for non-shows can be used for ways of predicting. There are many reasons for not showing up to an

appointment these include forgetting the appointment, other competing priorities or conflicts, and the patient's health status. To find a prediction model that will be better for every dataset, the four prediction algorithms are Artificial Neural Network, Gradient Boosting, Logistic Regression and Random Forest. These algorithms were chosen because they achieved the best results in previous no-show research. In the case of the prediction algorithms, none of them stands out but the one with more consistent results overall was Gradient Boosting. The author concludes that these predictions can help but are not still strong enough as a standalone strategy and should be combined with other scheduling strategies.

To conclude, various types of machine learning were used to predict the queue or waiting time. Among the machine learning method that is used are Linear Regression, Random Forest, Support Vector Machine (SVM) and Neural Network.

#### 2.3.1 Linear Regression

Linear regression is the simplest and most classic linear method for regression [21]. It is a mathematical approach used to perform predictive analysis when the prediction results can be quantified and modelled in relation to multiple input variables. It is a method to identify linear correlations between dependent and independent variables through analysing and modelling data [22]. Thus, this method would thus model relationships between dependent variables and independent variables from the analysis and learning to the current training result. Linear Regression is classified into three types: Simple Linear Regression, Multiple Linear Regression, and Polynomial Regression. However, in this project, simple linear regression will be considered. Simple Linear Regression is a case model with a

single independent variable. Simple Linear regression defines the dependence of the variable. The equation for simple linear regression as shown in equation 2.1.





$$Y_{i} = \beta_{0} + \beta_{1}X_{i} + \varepsilon_{i}$$
(2.1)

From equation 2.1, where  $X_i$  is the independent variable and  $Y_i$  is the dependent variable. The slope of the line is  $\beta_1$ . The intercept is  $\beta_0$  and the random error for the value is  $\varepsilon_i$ . Simple regression distinguishes the influence of independent variables from the interaction of dependent variables.

Simple Linear Regression was classified into 3 types of categories as shown below. Each category will explain the relationship among the variables.



Figure 2.5: Type of relationship continued



2.3.1.1 Visualising and a fitting a Linear Regression model

Statisticians say that a regression model fits the data well if the differences between the observations and the predicted values are small and unbiased. Unbiased in this context means that the fitted values are not systematically too high or too low anywhere in the observation space.

Generally, it is said that the range of R2 is 0 to 1. A value of 0 indicates that the dependent variable cannot be explained by the independent variable at all. Then, a value of 1 indicates that the dependent variable can be perfectly explained without error by the independent variable.



Figure 2.7: Relationship between the observed value and fitted value

#### 2.3.2 Random Forest

Random forest is a supervised machine learning algorithm used widely in Regression and Classification problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. Every tree is dependent on random vectors sampled independently, with similar distribution to every other tree in the random forest. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.



Another most widely used state-of-the-art machine learning technique is Support Vector Machine (SVM). Normally, SVM is leveraged for classification problems, **UNVERSITIEKNIKAL MALAYSIA MELAKA** constructing a hyperplane where the distance between two classes of data points is at its maximum. The hyperplane is the decision margin that separates the classes of data points (e.g., oranges vs. apples) on either side of the plane, as shown in Figure 2.9. Hyperplanes are decision margins that help classify the data points. The boundaries are drawn so that the distance between the margin and the classes is maximum and minimises the classification error [23].



Figure 2.9: Two classes of data point separated by a hyperplane

Given the independent and identically training samples described as  $s = \{(x1,y1), (x2,y2), ..., (xn,yn)\}$  where xi and yi are the input data and output of the classification respectively. The goal is to obtain hyperplane f for classification using Equation (2.2).  $y = f(x,w) = \langle x,w \rangle + b$ (2.2)

Where  $w \in \mathbb{R}N$  and  $b \in \mathbb{R}$  are the weighting factors,  $f:\mathbb{R}N \to \{-1,1\}$ . SVM classifier can be used to diagnose faults since it can handle the non-linear behaviour data well.

#### 2.3.4 Neural Network

An artificial neural network learning algorithm or neural network is a computational learning system that uses a network of functions to understand and translate a data input of one form into the desired output, usually in another form. The concept of the artificial neural network was inspired by human biology and the

way neurons of the human brain function together to understand inputs from human senses. Neural network is just tools and approaches used in machine learning algorithms. The neural network itself may be used as a piece in many different machine learning algorithms to process complex data inputs into a space that computers can understand.



Linear regression is one of the techniques of regressions commonly used for prediction analysis due to its comparative simplicity. Sherif et.al [24] use a linear regression model to examine the potential accuracy of predicted waiting time based on the patient's arrival in the Emergency Department (ED). The result demonstrates that this technique has the ability to predict the waiting time accuracy with minimal Mean Absolute error (MAE) compared to actual waiting time. A linear regression model is believed to develop more accurate predictive models if the variables are chosen correctly.

Moreover, the linear regression model also was tested for waiting time for outpatients in this project research [25]. A study was conducted to determine the factor that contributes to the long waiting time; thus, the author used linear regression to determine which variable contributes to the factor of long waiting time to improve the patient waiting time. Besides, the linear regression model seems capable of a machine learning algorithm that gives the second-highest accuracy to predict waiting time in a long queue [16]. Apart from that, Curtis et.al [17] presented a linear regression model that obtained the highest accuracy score among other machine learning algorithms to predict patient waiting time in a healthcare scenario.

On the other hand, linear regression was used to develop a system for predicting rubber crop yield [11]. This analysis chooses four parameters for predicting rubber crop yield: "temperature, rainfall, humidity, and planted area". Various machine learning algorithms were tested, such as Random Forest, Decision Tree, Linear Regression, and Neural Network. The performance of each algorithm is evaluated with Mean Squared Error (MSE) and Mean Absolute Error (MAE). The experimental result shows that the linear regression model provides the most accurate prediction among other algorithms since it has the least Mean Absolute Error (MAE) value.

### **CHAPTER 3**

### **METHODOLOGY**



In this chapter, the overall plan to achieve the objectives for the waiting time prediction and diagnosis will be explained. The methods and techniques used in the project will be discussed in detail.

#### **3.1** Overview of the project

Referring to Figure 3.1 shows the flow chart of this project. Initially, the database collected from Klinik Pergigian Alor Gajah is used throughout this project. Patient records for the years 2018 and 2019 are chosen for this project because it is the

actual waiting time scenario for pre-covid. This project is use Python Programming with Scikit-learn, the machine learning library for Python language. It consists of various algorithms and supports scientific libraries and Python numerical, Seaborn and NumPy. Furthermore, Pandas library is used to import datasets, while Matplotlib is used for graphical plotting. Before proceeding with feature selection, the data must be cleaned. Its purpose is to identify and eliminate mistakes in the raw dataset that may have a negative impact on a prediction model. Then, before splitting the dataset, perform feature selection. Feature selection is the crucial step for data preprocessing. Then, the dataset is normalised so that the dataset's variables lie within a specific range before splitting into training and testing datasets. The machine learning model algorithms, Linear Regression and Random Forest were used in this project. R-Squared, or coefficient of determination, may need to be greater than 90% before comparing to another machine learning model. After that, the prediction models' performance is compared and evaluated by computing the regression metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The important parameters are also identified and proposed with the best prediction model. Lastly, a graphical user interface (GUI) for the waiting time prediction system was developed using Tkinter for easy access to the dental clinic.



**Figure 3.1: Flowchart of the project** 

#### **3.2 Project Implementation**

This project employed Python 3 software version 3.9 which is more in demand and includes a typing system than Python 2, is used outdated and older syntax for the print function. Python 3.9 is a pure script object-oriented language design programme that integrates the essence and design rules of several design languages by displaying the features of interactive connections and the type of explanation. This project also uses a machine learning model which is linear regression and a Graphical User Interface (GUI). Meanwhile, Microsoft Excel was used for data cleaning and splitting data into training and testing datasets

#### **3.3 Data collection**

Data collection is a vital part of the research. It is the process of collecting information on interesting variables in an established systematic manner that enables one to answer stated research questions, test hypotheses and assess results. Dataset from Klinik Pergigian Alor Gajah, Melaka, was used to estimate the waiting time based on the best parameter. Patient records for the years 2018 and 2019 are chosen for this work because it is the actual waiting time scenario for pre-covid. The important parameters are also identified and proposed with the best prediction model.

Klinik Pergigian Alor Gajah 2018.xlsx
 Klnik Pergigian Alor Gajah 2019.xlsx

**Figure 3.2: Patient records raw dataset** 

#### **3.4** Data cleaning

Data cleaning also is a critically important step before doing the feature selection. It is to identify and correct the errors in the raw dataset that may negatively impact a predictive model. There are many types of errors in a dataset, although some of the simplest errors include the columns that contain not much information and duplicated rows. For example, if our dataset is in a text-based format, we have to convert it into a numerical value so that our programs can be read clearly by the computer. If these errors cannot be fixed early, then our model will learn a bad pattern and gain errors. Data cleaning process was for the years 2018 and 2019, which consisted of row and column data (5000, 17).

#### **3.5 Feature selection**

The selected data were labelled with their respective types. The data types were divided into 8 categories. First, is the registration of patients that are attending the dental clinic. Second, is the appointment status of the patients. Third, how old are the patients present at the clinic. Fourth, the category of the patients is based on age. Fifth is the chance of the patients getting fast lane or not. Sixth, the type of treatments in a dental clinic. Seventh, is the waiting time for the patients to seek treatment from the doctor. Lastly, the estimated waiting time for the patient to get treatment. Different values of data for every parameter were collected to predict the time waiting in the queue. The labelled types with their respective conditions were tabulated in Table 3.1.

Types	Conditions
New Registration	0: No
	1: Yes
Appointment	0: No
	1: Yes
Age	The age of the patient
Category	0: Toddler
	1: School
	2: Adult
MALAYSIA	3: Pregnant
	4: Senior Citizen
AS TEN	5: Disabled people
Fast lane	0: No
	1: Yes
Type of treatments	0: Extraction
<b>UNIVERSITI TEKNIKAL</b>	1: Scaling
	2: Filling
	3: Cleaning
	4: X-ray
	5: Medical check-up
Waiting time	The waiting time already have in the dataset
Estimate waiting time in minutes	0: time < 30
	1: 30 < time < 60
	2: 60 < time < 90
	3: 90 < time < 120

### Table 3.1: List of data types

#### 3.6 Split data

In a basic two-part data split, the training data set is used to train and develop models. Training sets are commonly used to estimate different parameters or to compare different model performances. The testing data set is used after the training is done. Therefore, the overall collected data were split into training data as much as 80% and testing data as much as 20%. The technique of data partitioning is shown in Figure 3.3, and the cycle was repeated for all the 5000 data for the years 2018 and 2019.



80% of the overall collected data was set as training data. Thus, from the overall 5000 data, 4000 of them were partitioned as training data and were used in training the linear regression model. Another 20% which was 1000 data was used to test the linear regression model.

#### 3.7 Train Linear Regression model

The given dataset was imported in Python to build a Linear Regression model. Before that, we need to understand the model work. There are many different machine learning algorithms. Each algorithm has advantages and disadvantages in terms of errors and its performance, so the algorithm that we use depends on kind of the problem that we trying to solve for the input data.

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an independent variable, and the other is considered to be a dependent variable.

#### 3.7.1 Making Predictions from Linear Regression models

The most important thing before creating a script is must import the Python library first. A Python library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes Python programming simpler and more convenient for the programmer.

UNIVPython library KNIKAL	IALAYSIA MELANA
Pandas	Used for data cleaning and analysis
Seaborn	Provides a high-level interface for drawing
	attractive and informative statistical graphics
Train test split	Used for splitting data arrays into training
	data and testing data
Metrics	Used for measures of quantitative
	assessment commonly used for assessing,
	comparing and tracking performance model

# Table 3.2: Python library and function

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

#### **Figure 3.4: Import Python library**

After that, the describe function is used for calculating some statistical data like how much data is counted in the data frame, mean, standard deviation, percentile, minimum and maximum value. It analyses both numeric and object series and also the data frame column sets of mixed data types.

	H.					
c.	linic	_dataset.de	scribe()			
		an-				
	.1.	Category	Treatment	Treatment_average	Total_patient	Waiting_time
c	ount	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
I	mean	2.921000	2.728000	24.220000	3.526000	3.556000
	std	1.030035	1.508135	7.127535	1.587397	1.596049
	min	1.000000	1.000000	10.000000	1.000000	1.000000
	25%	2.000000	2.000000	20.000000	2.000000	2.000000
	50%	3.000000	3.000000	20.000000	4.000000	4.000000
	75%	3.000000	3.000000	30.000000	5.000000	5.000000
	max	5.000000	7.000000	40.000000	6.000000	6.000000

Figure 3.5: Clinic dataset describes

In the clinic dataset, there are few data in text-based. The machine cannot understand and read that data. Therefore, this dataset needs to convert from the string into an integer so that the computer can read the language.

#### # Replace characters into numerical values

clinic\_dataset.replace({'Category':{'Toddler':1,'School':2,'Adult':3,'Pregnant':4,'Senior Citizen':5}},inplace=True)
clinic\_dataset.replace({'Treatment':{'Extraction':1,'Scaling':2,'Filling':3,'Cleaning':4,'X-RAY':5,'Medical Check Up':6,'
clinic\_dataset.replace({'Waiting\_time':{'0<waiting<30':1,'30<waiting<60':2,'60<waiting<90':3,'90<waiting<120':4,'120<wait
</pre>

Figure 3.6: Replace characters with numerical values

This is a dataset after the conversion from the text-based into a numerical value. Three parameter value has been replaced such as category, treatment and waiting time. The value can be observed in Figure 3.7 below.

# UNIVERSITATE Character () AL MALAYSIA MELAKA

	Category	Treatment	Treatment_average	Total_patient	Waiting_time
0	3	4	20	1	1
1	3	4	20	1	1
2	3	4	20	1	1
3	3	7	40	1	1
4	3	7	40	1	1

Figure 3.7: Clinic dataset

Before building the linear regression model, we are choosing 80 % of the dataset to be used as our training data. Basically, we just want to take the other 20% and save that for our testing data.

In [16]: M X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size = 0.2, random\_state=2)
lin\_reg\_model = LinearRegression()
lin\_reg\_model.fit(X\_train,Y\_train)

Out[16]: LinearRegression()

#### Figure 3.8: Build the linear regression model

After building the model, we evaluate the performance of linear regression which were the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Rsquare is a measure of how well changes in the dependent variable can be predicted by changes in the independent variables. These regression metric values have been obtained for both training and testing data.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

```
training_data_prediction = lin_reg_model.predict(X_train)
test_data_prediction = lin_reg_model.predict(X_test)
print("Training MAE = ", (mean_absolute_error(Y_train,training_data_prediction)))
print("Training RMSE = ", np.sqrt(mean_squared_error(Y_train,training_data_prediction)))
print("Training R2 = " , r2_score(Y_train,training_data_prediction))
print('\n')
print("Testing MAE = ", (mean_absolute_error(Y_test,test_data_prediction)))
print("Testing RMSE = ", np.sqrt(mean_squared_error(Y_test,test_data_prediction)))
print("Testing RMSE = ", np.sqrt(mean_squared_error(Y_test,test_data_prediction)))
print("Testing R2 = " , r2_score(Y_test,test_data_prediction)))
Training MAE = 0.12239302938794495
Training RMSE = 0.3087174433810152
Training R2 = 0.9631964551097814
```

```
Testing MAE = 0.10464068124842005
Testing RMSE = 0.2711712210164143
Testing R2 = 0.9687090080393477
```

**Figure 3.9: Evaluation regression metric values** 

In this Jupyter Notebook, we try to predict the based on the selected parameter in

the dataset. The result of the predicted value was obtained from Figure 3.10.

To prove the result of the predicted value in Jupyter Notebook, we have to

aug

compare it with the theoretical calculation. The result of the theoretical calculation value was obtained from Figure 3.11 below. In conclusion, we got the same predicted value after comparing the results for both Figure 3.10 and Figure 3.11.

	<pre># Predict result with labelling the independent variables lin_reg_model.predict([[1,2,3]])</pre>
	C:\Users\HP\AppData\Local\Programs\Python\Python39\lib\sit feature names, but LinearRegression was fitted with featur warnings.warn(
$\left( \right)$	array([3.12853751])

Figure 3.10: Predicted value



**Figure 3.11: Theoretical calculation value** 

The parameter was selected which were the category, type of treatments and total of the patient as independent variable and the predicted waiting time as a dependent variable as shown in the figure below.



**Figure 3.12: Independent variables** 

print	(Y)				
0	1				
1	1				
2	1				
3	1				
4	1				
	••				
4995	4				
4996	4				
4997	4				
4998	4				
4999	4				
Name:	Waiting_time,	Length:	5000,	dtype:	int64

Figure 3.13: Dependent variable

Figure 3.14 represents the plot of the actual waiting time versus the predicted waiting time using the Linear Regression model. The actual waiting time is labelled "Y train" in Jupyter Notebook, while the predicted waiting time is labelled "training data prediction."

```
plt.scatter(Y_train, training_data_prediction)
plt.xlabel("Actual Waiting Time")
plt.ylabel("Predicted Waiting Time")
plt.title(" Actual Waiting Time vs Predicted Waiting Time")
plt.show()
```



Figure 3.14: Build a Linear Regression model

#### 3.7.2 Root Mean Square Error

Root Mean Square Error (RMSE) is a quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the average of squared differences between prediction and actual observation.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (y_j - \hat{y}_j)^2}$$
(3.1)

According to the equation in 3.1,  $y_j$  is the predicted value and  $\hat{y}_j$  is the actual value of the observation. Finally, n refers to the total number of observations.

#### 3.7.3 Mean Absolute Error

Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

$$MAE = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_j|$$
(3.2)

According to the equation in 3.2,  $y_j$  is the predicted value and  $\hat{y}_j$  is the actual value of the observation. Finally, n refers to the total number of observations.

#### 3.8 Comparison of Machine Learning Models

The Linear Regression Model was compared with other Machine Learning models which was the Random Forest Regression. The Random Forest Regression model was developed using Jupyter Notebook and other machine learning libraries based on the same training data and tested using the same testing data. This is to ensure the comparison between each machine learning is unbiased.

The categories used for the comparison between the two machine learning models were the RMSE, MAE and R-Squared of the result obtained from models. These categories' evaluation metrics are the important elements in analysing the efficacy of the machine learning models. The results of these evaluation metrics were obtained and analysed in the next chapter.

# 3.9 Development of Graphical User Interface (GUI) for Queue System for Waiting Time Prediction

Graphical User Interface (GUI) is a form interface that allows users to interact with electronic devices through graphical icons. It is used to communicate with the computer by moving a pointer around on a screen.

#### UNIVERSITI TEKNIKAL MALAYSIA MELAKA

A queue system for waiting for time prediction based on Linear Regression was developed using Jupyter Notebook and other machine learning libraries. The designed queue system for waiting time prediction GUI contained a single interface which to predict queue waiting time. The interface of GUI is shown in Figure 3.15. It allowed the user to enter a value based on category and the button used to predict the waiting time for the patient to get served by the doctor.

🆉 Syamsul   Final Year Project	_		×
Queue System For Waiting Time Prediction Using Lir	near Regress	ion	
Category Treatment Total patient			
Category : 1) Toddler 2) School 3) Adult 4) Pregnant Treatment : 1) Extraction 2) Scaling 3) Filling 4) Cleani Predict	5) Senior C ing 5) X-RA	itizen Y	
Waiting time : 1) 0< waiting<30 2) 30< waiting<60 3) 60< waiting	ng<90 4) 90	<waiting< td=""><td>&lt;120</td></waiting<>	<120
Figure 3.15: GUI of Queue System for Waitin	g Time P	redicti	on
سيتي تيكنيكل مليسيا ملاك	ينوس	١	

For the development of the queue system for waiting time prediction GUI, it was ensured that the design interface was simple and user-friendly. It was an important element which allowed anyone to use GUI without requiring deep knowledge such as data cleaning, feature selection, dataset splitting, and machine learning model training. The results obtained were displayed at the bottom section of the GUI using the Linear Regression model.

### **CHAPTER 4**

### **RESULTS AND DISCUSSION**



This chapter will discuss the result obtained from the previous methodology section. All the collected results will be in form of a table, graph or picture in order to demonstrate the efficacy of the machine learning model. The evaluation metrics of each model will be explained and analysed in detail. The findings of this section will be used to review the objectives of this project.

#### 4.1 Linear Regression Model

The Linear Regression Model for the queue system for waiting time prediction was successfully trained and developed in Jupyter Notebook. Figure 4.1 shows the developed linear regression trained by using a machine learning algorithm. Feature selection improves the machine learning process and increases the predictive power of machine learning algorithms by selecting the most important variables and eliminating redundant and irrelevant parameters. This step is important to avoid the over-fitting or under-fitting of the machine learning model. From the linear regression, 3 independent variables were selected which were the category, treatment and total of patients. The waiting time was selected as a dependent variable. As a result, not all the parameters were selected to generate the Linear Regression, instead only the best parameters are chosen.



**Figure 4.1: Linear Regression Model** 

Figure 4.1 above shows the relationship between the actual waiting time and the predicted waiting time. The value of coefficient and value of intercept are already

got from Figure 3.11 in Jupyter Notebook. The graph showed that the actual and predicted waiting times range from 1 to 6, as classified in Table 1. For ranges 1, 2, 3, and 6, it is possible to conclude that the dataset can be predicted nearly 80% accurately. Next, for range 4, the dataset can be predicted nearly 20% of the time. Finally, only 40% of the range 5 can be predicted. To summarise, it is still on a good performance model with an R-Squared of more than 95% that fits well with the data.

#### 4.1.1 Performance of Linear Regression Model

Table 4.1 shows the overall performance of the Linear Regression model. The model was able to predict a queue system for waiting time for both training and testing data. The obtained Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the training data of this model were almost the same compared to the testing data. RMSE and MAE are important to evaluate the performance and exactness of the generated machine learning model. The results of these metrics regression have been obtained from Figure 3.9.

Evaluation metrics	Training data	Testing data
Root Mean Square Error	0.31	0.27
Mean Absolute Error	0.12	0.10
R-Squared	0.96	0.97

 Table 4.1: Linear Regression of regression metrics

#### 4.1.2 Comparison with Random Forest Regression Model

Table 4.2 shows the overall performance of Random Forest Regression Model. The obtained Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the training data of this model were almost the same compared to the testing data. Both RMSE and MAE for training data and testing data for this model were slightly lower compared to the Linear Regression Model.

Evaluation metricsTraining dataTesting dataRoot Mean Square Error0.340.30Mean Absolute Error0.160.14R-Squared0.930.94

**Table 4.2: Random Forest of regression metrics** 

#### 4.2 Queue System for Waiting Time Prediction using Developed Graphical

User Interface (GUI)

The Linear Regression Model was integrated into the developed GUI for the UNIVERSITITEKNIKAL MALAYSIA MELAKA queue system for waiting time prediction. A single interface was developed in this project. It allows the user to predict queue waiting time to get treatment from the doctor.

#### 4.2.1 Testing of an interface of Queue System for Waiting Time Prediction

Figure 4.2 shows how the GUI works with patient data, given in values. The GUI successfully predict the waiting time for each patient after inserting the patient data. The independent variable has been listed for guidance on how to use it. For example,

the category which was toddler, school, adult, pregnant and senior citizen was labelled with a numerical value. Besides that, the treatment is also has been labelled in numerical for easy to predict the waiting time. As the result, the predicted was displayed in numbers in minutes.

🧳 Syamsul   Final Year Project	—		$\times$
Queue System For Waiting Time Prediction Using Lir	near Regressi	on	
Category 3 Treatment 2 Total patient 2			
Category : 1) Toddler 2) School 3) Adult 4) Pregnant Treatment : 1) Extraction 2) Scaling 3) Filling 4) Clean Predict Predict {Predicted Waiting Time : } 2	5) Senior Cit ing 5) X-RAY	izen	
Waiting time: 1) 0 waiting <30 2) 30 < waiting <60 3) 60 < waiting	ng<90_4) 90<	waiting<	120
UNIVERSITI TEKNIKAL MALAYSIA Figure 4.2: Testing of GUI for pre	MELAK	A	

Figure 4.2 showed the predicted waiting time in minutes after inserting the patient data. The results showed that the predicted waiting time is 2, meaning that the patient must wait between 30 until 60 minutes before getting treatment from the doctor. The patient still has time for other things like eating breakfast, using the toilet or other than that.

#### 4.3 Environmental and Sustainability

This project is made environment friendly which is the project does not need any paper ticket or form to be filled up during the registration. Therefore, the usage of paper will be reduced. Besides that, this system does not use any dangerous things that can be affected by the user. In conclusion, this project brings benefits and is safe to use in the future

According to sustainability, the project is not a one-time-use application. The patient will use it every time they visit the dental clinic. Next, this project can contribute to an organization in a long-term application and reduce maintenance costs. This project also contributes to Sustainable Development Goals (SDG) target number 7, which is affordable and clean energy.

#### 4.4 Chapter summary

From the overall result obtained, the developed Linear Regression model has a good performance as its RMSE and MAE have the lowest value which has the least error and is a good model in both training and testing data. It was able to predict the waiting time of each patient using the Linear Regression model. In summary, the Linear Regression using machine learning was suitable to predict queue for waiting time. The value of R-Squared for both training and testing data was above 95% which is the variation in the output can be explained by the input variable.

The other machine learning model, Random Forest was the second model where its RMSE and MAE value were a bit higher than the Linear Regression model. The result of both training and testing data is also considered a good model which has the least error. Besides that, the value of R-Squared for both training and testing data was above 90%, almost having a perfect fit line.

In summary, Linear Regression is a better model than Random Forest since it has less RMSE and MAE and a higher R-squared or coefficient of determination. The development of GUI using Linear Regression for waiting time prediction is important as it enables the user to predict the waiting time required without studying in more depth like required to preprocess, train the dataset and train the machine learning model. A waiting time can be predicted by using the GUI system. This system can contribute to an organization by increasing components' lifetime and reducing maintenance costs. Besides that, this system protects the environment by reducing energy generation because the waste of energy situation is eliminated.



### **CHAPTER 5**

### **CONCLUSION AND FUTURE WORKS**



#### 5.1 Conclusion

The purpose of this project is to predict queue waiting time by employing the machine learning technique. The first objective of this project is to develop a time prediction system using Linear Regression and Random Forest, a machine learning algorithm. Two machine learning methods were compared and evaluated. A Linear Regression model for queue waiting time was successfully trained and developed in Jupyter Notebook using Python language. This model was able to predict the queue waiting time by selected parameters which were the category, treatment and total of

the patient as an independent variable and the waiting time as a dependent variable. The performance of this model was evaluated using RMSE and MAE regression metrics. As a result, Linear Regression outperformed other machine learning in predicting the waiting time scenario.

The second objective of this project is to design a Graphical User Interface (GUI) that much easier for the user to interact with it. A GUI with a single interface is successfully developed using Tkinter Framework. The trained Linear Regression model for queue waiting time is integrated into the GUI system to predict the independent variables and get the result of the queue waiting time which is a dependent variable. The design of the GUI interface was simple and user-friendly. It was an important element which allowed anyone to use GUI without requiring deep knowledge.

In a conclusion, this project is fully completed as all the objectives are achieved perfectly.

### UNIVERSITI TEKNIKAL MALAYSIA MELAKA

#### 5.2 Future Works

Machine Learning (ML) is already lending a hand in diverse situations in healthcare. ML in healthcare helps to analyze thousands of different data and suggest outcomes. It also increases the ability of healthcare professionals to better understand the day-to-day patterns.

This project was successful in predicting the patient's waiting time. However, this project can predict the waiting time with only one doctor. By increasing the total number of doctors, this system can become more effective, and also the sort of treatment is more available. This is because some doctors may have expertise in a specific treatment, which might save the service time. Increasing the size of the dataset will improve prediction accuracy. This improvement can help patients avoid delays.

Aside from that, each treatment has a different range of service time based on the doctor and the situation. To get an accurate waiting time prediction, visit a dental clinic and record each time. It was quite difficult, but I believe the results will satisfy my expectation. It could be an error, but it would be the least error before this project. Following that, this project can be tested at a dental clinic in real-time. In the future, this approach will make the patient satisfied with the facilities at the government dental clinic.

Finally, by implementing this project, it will be possible to prevent congestion in the dental clinic. With the deployment of the system using the Jetson nano microcontroller to make the system a standalone system, the patient now knows the expected waiting time. When the dental clinic is not crowded, the air conditioner is used efficiently. As a result, thermal comfort can be obtained while also providing convenience for staff and patients. If this prediction project is not developed, the patient cannot estimate the time and will stay in the clinic until called for the treatment. The situation in the dental clinic will be uncomfortable for the staff and patients. In conclusion, this project prediction brings benefits to other users.

### REFERENCES

- M. M. Kembe, E. S. Onah, and S. Iorkegh, "A Study of Waiting And Service Costs of A Multi- Server Queuing Model In A Specialist Hospital," *Ijacst*, vol. 1, no. 8, pp. 19–23, 2012.
- [2] Ministry of Health Malaysia, "Garispanduan Perlaksanaan Pengurusan Pemantauan Waktu Menunggu di Jabatan Pesakit Luar (JPL) dan Klinik Kesihatan," pp. 6–25, 2008.
- [3] M. A. A. Hasin, R. Seeluangsawat, and M. A. Shareef, "Statistical measures of customer satisfaction for health care quality assurance: A case study," *Int. J. Health Care Qual. Assur.*, vol. 14, no. 1, pp. 6–13, 2001, doi: 10.1108/09526860110366214.
- [4] A. B. Azraii, K. N. Kamaruddin, and F. Ariffin, "An assessment of patient waiting and consultation time in a primary healthcare clinic," *Malaysian Fam. Physician*, vol. 12, no. 1, pp. 14–21, 2017.
- [5] M. Ngorsed and P. Suesaowaluk, "Hospital service queue management system with wireless approach," *Lect. Notes Electr. Eng.*, vol. 375, pp. 627–

- [6] M. H. Naim, R. R. Raja Ikram, L. Salehudin, W. N. Abd Rashid, M. K. Abd Gani and N. H. Ismail, "The design of mobile push notification for improving queuing system in electronic health services," vol. 2018, no. November, pp. 149–150, 2018.
- [7] M. S. Helmy, S. I. Ibrahim, and L. Bala Subramaniam, "Wearable Wristbands for E-Queue Hospital," Melaka, 2021.
- [8] V. Nath and S. E. Levinson, *Machine learning*, vol. 0, no. 9783319056050.
   2014. MALAYSIA
- [9] R. P. Satya Hermanto, Suharjito, Diana, and A. Nugroho, "Waiting-Time Estimation in Bank Customer Queues using RPROP Neural Networks," *Procedia Comput. Sci.*, vol. 135, pp. 35–42, 2018, doi: 10.1016/j.procs.2018.08.147.
- [10] N. A. Sulaiman, M. P. Abdullah, H. Abdullah, M. N. S. Zainudin, and A. M. Yusop, "Fault detection for air conditioning system using machine learning," *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, pp. 109–116, 2020, doi: 10.11591/ijai.v9.i1.pp109-116.
- [11] H. J. Pang, M. N. S. Zainudin, F. Kejuruteraan, E. Dan, and K. Komputer, "Analysis on prediction of rubber crop production using machine learning."
  [Online]. Available: http://www.fao.org/3/ca6514en/ca6514en.pdf.
- [12] A. I. Kyritsis and M. Deriaz, "A machine learning approach to waiting time prediction in queueing scenarios," in *Proceedings 2019 2nd International*

Conference on Artificial Intelligence for Industries, AI4I 2019, Sep. 2019, pp. 17–21, doi: 10.1109/AI4I46381.2019.00013.

- [13] P. Anussornnitisarn and V. Limlawan, "Design of advanced queue system using artificial neural network for waiting time prediction," *Revistaespacios*, vol. 41, no. 40, pp. 111–124, 2020.
- [14] R. P. Satya Hermanto, Suharjito, Diana, and A. Nugroho, "Waiting-Time Estimation in Bank Customer Queues using RPROP Neural Networks," *Procedia Comput. Sci.*, vol. 135, pp. 35–42, 2018, doi: 10.1016/j.procs.2018.08.147.
- [15] R. S. Carvalho, R. N. Carvalho, G. N. Ramos, and R. N. Mour<sup>o</sup>o, "Predicting waiting time overflow on bank teller queues," *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017*, vol. 2017-Decem, pp. 842–847, 2017, doi: 10.1109/ICMLA.2017.00-51.
- [16] D. Gomes, R. H. Nabil, and K. Nur, "Banking queue waiting time prediction based on predicted service time using support vector regression," in *Proceedings of International Conference on Computation, Automation and Knowledge Management, ICCAKM 2020*, Jan. 2020, pp. 145–149, doi: 10.1109/ICCAKM46823.2020.9051490.
- [17] C. Curtis, C. Liu, T. J. Bollerman, and O. S. Pianykh, "Machine Learning for Predicting Patient Wait Times and Appointment Delays," *J. Am. Coll. Radiol.*, vol. 15, no. 9, pp. 1310–1316, Sep. 2018, doi: 10.1016/j.jacr.2017.08.021.
- [18] R. Md Noor et al., "PREDICT ARRIVAL TIME BY USING MACHINE

- [19] Y. Sanit-In and K. R. Saikaew, "Prediction of waiting time in one-stop service," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 3, pp. 322–327, Jun. 2019, doi: 10.18178/ijmlc.2019.9.3.805.
- [20] R. Almeida, N. A. Silva, and A. Vasconcelos, "A machine learning approach for real time prediction of last minute medical appointments no-shows," in *HEALTHINF 2021 - 14th International Conference on Health Informatics; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, 2021, pp. 328–336, doi: 10.5220/0010221903280336.
- [21] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," J. Appl. Sci. Technol. Trends, vol. 1, no. 4, pp. 140–147, 2020, doi: 10.38094/jastt1457.
- [22] H. Il Lim, "A linear regression approach to modeling software characteristics for classifying similar software," *Proc. Int. Comput. Softw. Appl. Conf.*, vol. 1, pp. 942–943, 2019, doi: 10.1109/COMPSAC.2019.00152.
- [23] M. Batta, "Machine Learning Algorithms A Review," Int. J. Sci. Res. (IJ, vol. 9, no. 1, pp. 381-undefined, 2020, doi: 10.21275/ART20203995.
- [24] S. A. K. Hemaya and T. E. Locker, "How accurate are predicted waiting times, determined upon a patient's arrival in the Emergency Department?," *Emerg. Med. J.*, vol. 29, no. 4, pp. 316–318, 2012, doi:

10.1136/emj.2010.106534.

S. T. T. Nguyen *et al.*, "Waiting time in the outpatient clinic at a national hospital in Vietnam," *Nagoya J. Med. Sci.*, vol. 80, no. 2, pp. 227–239, 2018, doi: 10.18999/nagjms.80.2.227.

