



Faculty of Computer Science and Information Technology

Analysis on Dengue's Seasonality in Malaysia

Tan Jia Er

58069

Bachelor of Computer Science with Honors (Computational Science)

2019/2020

Analysis on Dengue's Seasonality in Malaysia

TAN JIA ER

This project is submitted in partial fulfilment of the
requirements for the degree of Bachelor of
Computer Science with Honors

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2019/2020

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE ANALYSIS ON DENGUE'S SEASONALITY IN MALAYSIA

ACADEMIC SESSION: 2019/2020

TAN JIA ER

(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 in order 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

Validated by

Jas.
(AUTHOR'S SIGNATURE)

Jme
(SUPERVISOR'S SIGNATURE)

Permanent Address

495 Parit Sikun, Jalan Kluang,
86400 Parit Raja, Johor.

Date: 3/8/2020

Date: 3/8/2020

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

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

DECLARATION

I hereby declare that the project entitled “Analysis on Dengue’s Seasonality in Malaysia” submitted by me to Faculty of Computer Science and Information Technology, University of Malaysia Sarawak (UNIMAS) is a record of an original work done by me under guidance of Associate Professor Dr. Jane Labadin.



TAN JIA ER

ACKNOWLEDGEMENT

I would like to thank my supervisor, Associate Professor Dr. Jane Labadin and also my examiner, Mr. Terrin Lim for the feedback of proposal, advice, guidance and encouragement throughout my final year project.

I would also like to thanks my friends who are giving me precious opinions and help throughout my final year project.

TABLE OF CONTENTS

DECLARATION.....	i
ACKNOWLEDGEMENT.....	ii
LIST OF FIGURES	v
LIST OF TABLES	vii
LIST OF EQUATION	viii
ABSTRACT	ix
ABSTRAK	x
CHAPTER 1 INTRODUCTION.....	1
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Objectives.....	3
1.4 Methodology	3
1.5 Scope	5
1.6 Significance of Project	5
1.7 Project Schedule.....	6
1.8 Report Organization	7
1.9 Project Outcome.....	8
1.10 Summary	8
CHAPTER 2.....	9
2.1 Introduction	9
2.2 Approaches of Dengue Seasonality Studies.....	9
2.2.1 Machine Learning.....	10
2.2.2 Autoregressive Integrated Moving Average (ARIMA).....	12
2.2.3 Compartmental Model.....	16
2.4 Comparison between review papers.....	19
2.5 Summary	21
CHAPTER 3.....	23
3.1 Introduction	23
3.2 Data processing	24
3.3 Model Formulation.....	26
3.3.1 Pearson Correlation	28
3.4 ARIMA.....	34

3.4.1 Steps for building ARIMA model	36
3.5 Summary	37
CHAPTER 4.....	38
4.1 Introduction	38
4.2 Model Analysis	39
4.3 ARIMA model.....	41
4.3.1 Model Analysis.....	44
4.3.2 Model Summary	45
4.4 Summary	45
CHAPTER 5.....	46
5.1 Introduction	46
5.2 Conclusion.....	46
5.3 Limitations	47
5.4 Future Works.....	48
REFERENCES	49
APPENDIX.....	50

LIST OF FIGURES

Figure 1.1: Summary on Methodology of Modelling Process.....	4
Figure 1.2: Gantt chart of FYP 1 and FYP 2	7
Figure 2.1: Types of machine learning (MathWorks)	10
Figure 2.2: Flowchart of ARIMA	13
Figure 2.3: SIR Model	17
Figure 3.1: Research Framework.....	23
Figure 3.3: Total number of dengue cases in year 2010-2017.....	26
Figure 3.3A: Minimum temperature and total number of dengue cases in year 2010-2017 ..	27
Figure 3.3B: Average relative humidity and total number of dengue cases in year 2010-2017	27
Figure 3.3C: Total rainfall and total number of dengue cases in year 2010-2017	28
Figure 3.2A: Scatterplot of number of dengue cases against minimum temperature.....	30
Figure 3.2B: Scatterplot of number of dengue cases against average relative humidity.....	30
Figure 3.2C: Scatterplot of number of dengue cases against total rainfall.....	31
Figure 3.3: The Pearson's product-moment correlation between the number of dengue cases and minimum temperature, T_{min}	32
Figure 3.4: The Pearson's product-moment correlation between the number of dengue cases and average relative humidity, H	32
Figure 3.5: The Pearson's product-moment correlation between the number of dengue cases and total rainfall, PP	33
Figure 4.1: School holidays correspond to total number of dengue cases in year 2010-2017	39

Figure 4.2: The Pearson’s product-moment correlation between the number of dengue cases and school holidays.....	40
Figure 4.3: Augmented Dickey-Fuller Test for the number of dengue cases generated by R	41
Figure 4.4: Augmented Dickey-Fuller Test for the number of dengue cases after performing 1 time of differencing generated by R	42
Figure 4.5A: The PACF graph for the total number of dengue cases	42
Figure 4.5B: The ACF graph for the total number of dengue cases	42
Figure 4.6: Forecasting graph of the number of dengue cases with ARIMA (3,1,4)	43
Figure 4.7: The total number of dengue cases statistical information generate by R with ARIMA (3,1,4).....	44

LIST OF TABLES

Table 2.1: Related dengue studies with climate variables applying modelling approach	19
Table 4.1: Comparison of tested models.....	44
Table A1: Number of dengue cases occurred in Selangor corresponds to the epidemiological week from year 2010 to year 2017	50
Table A3: Climate data for Selangor in year 2011	53
Table A4: Climate data for Selangor in year 2012	54
Table A5: Climate data for Selangor in year 2013	56
Table A6: Climate data for Selangor in year 2014	57
Table A7: Climate data for Selangor in year 2015	58
Table A8: Climate data for Selangor in year 2016	60
Table A9: Climate data for Selangor in year 2017	61
Table A10: Malaysia school holidays and term dates from Year 2010 to Year 2017	63
Table A11: Number of dengue cases relate to school holidays (Binary variable: 0-No; 1-Yes)	64

LIST OF EQUATION

Equation 1: $Y_t = \theta + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} + a_t$	35
---	-----------

ABSTRACT

In Malaysia, there is currently not have enough study which discuss about the climatic factors and also dengue cases per epidemiological week. The purpose of this study was to investigate the correlation between the climatic factors such as minimum temperature, average relative humidity and total amount of rainfall with the number of dengue cases recorded per epidemiological week. Selangor is chosen as the study area because it achieved the highest number of dengue cases occurred within the period of year 2010 to year 2017. The number of dengue cases are retrieved from Malaysia's Open Data Portal while the weather data are retrieved from website of Tutiempo. Pearson's product-moment correlation test was done between the number of dengue cases and the climatic factors. The result obtained shows that there exists a weak relationship between the climatic factors and the number of dengue cases. Besides, ARIMA model is used to predict the future dengue cases happens. The ARIMA model can be used to optimize and predict the dengue occurrence by prediction on the climate effect on dengue cases in Malaysia.

ABSTRAK

Pada masa ini, masih tidak banyak kajian yang membincangkan tentang faktor iklim dan juga kes demam denggi mengikut minggu epidemiologi. Tujuan kajian ini adalah untuk mengkaji korelasi antara faktor-faktor iklim iaitu suhu minimum, kelembapan relatif purata dan jumlah hujan dengan bilangan kes denggi yang dicatatkan mengikut minggu epidemiologi. Selangor dipilih sebagai kawasan kajian kerana ia merupakan kawasan di antaranya Malaysia yang mencatat kes demam denggi yang tertinggi dalam tempoh tahun 2010 hingga tahun 2017. Jumlah kes denggi mengikut minggu epidemiologi adalah diambil dari Portal Data Terbuka Malaysia manakala data cuaca adalah diambil dari laman web Tutiempo. Ujian “Pearson product-moment correlation” dilakukan antara kes denggi dan faktor iklim. Hasilnya menunjukkan terdapat hubungan yang kurang signifikan antara faktor iklim dan jumlah kes denggi. Model ARIMA digunakan untuk meramalkan kes denggi yang berlaku pada masa depan. Selain itu, model ARIMA juga boleh digunakan untuk mengoptimumkan dan meramalkan pencegahan denggi dengan ramalan tentang kesan iklim terhadap kes demam denggi di Malaysia.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Dengue was classified by The World Health Organization (WHO) in 2012, as one of the most rapidly spreading mosquito-borne viral disease in the world. It is caused by dengue viruses (DENV 1-4), which are transmitted to human hosts through the bites of a striped *Aedes aegypti* mosquitoes that has previously bitten an infected person (Hii et al., 2016). Most of the *Aedes* mosquitoes can be found throughout the tropical and subtropical areas, and they have adapted to cohabiting with humans in both the urban and the rural environment. The dynamics of dengue transmission are influenced by multiple complex risk factors which are host immunity, vector capacity, circulating DENV, weather or climate, dengue control capacity, and population movement.

In Malaysia, dengue is one of the most important vector-borne diseases in terms of human morbidity and mortality. Malaysia is a tropical country whereby there is no four seasons, with a relatively high temperature and high relative humidity and copious rainfall throughout the year. All these conditions are conducive for the growth and survival rate of *Aedes* mosquitoes. Climate variables can affect the life-cycles, survival rates and biting rates of mosquitoes, as well as the incubation period of DENV.

Dengue has been endemic in Malaysia since the 1970s, with increasing intensity and magnitude of outbreaks in recent decades. In the past decades, studies have shown evidence of a relationship between climate and dengue and of the feasibility of using climate data to predict dengue outbreak (Hii et al., 2016). The climate is characterized by high average temperatures

and rainfall, with only small differences in temperature reported throughout the year. High rainfall patterns follow the monsoon winds, which occur between November and March and from June to September. The seasonality of dengue can be decomposed into two components which are climate-driven seasonality and non-climate driven seasonality. Non-climate trends included short and long-term trends, and also involves the effect of school sessions, public holidays, and weekday-weekend pattern on these diagnoses. Climate-driven trends included local climate variables impacts such as rainfall and humidity on dengue fever diagnoses.

These studies are important for the advanced development of seasonality-based modelling that could pave the way for an early warning tools for dengue outbreak. An effective early warning system can enhance the dengue control and also reduces the health and economic burden of dengue in the population. However, dengue disease is likely to remain endemic for a long time and will remains as a major public health concern in Malaysia. Therefore, this research aims to determine the dengue's outbreak season as well as formulate the dengue outbreak incorporating the seasonality model.

1.2 Problem Statement

The reported cases of dengue in Malaysia has increased dramatically in recent years. As the number of dengue cases increases, a lot of studies and research have been done to investigate the factor and impact of dengue outbreak. World Health Organization (WHO) recommends the best way to prevent dengue endemic is to provide knowledge about dengue and also preventive measures to people. Despite of the efforts to prevent dengue virus, prevalence of dengue among population is still on the rise.

Dengue has become a major public health concern for over half of the world's population owing to its detrimental effects on people's health and is known as one of the leading causes of hospitalization and death. Therefore, it is important to make prediction on the dengue seasonality for precautions purpose during that peak season to minimize the loss.

The seasonal pattern of dengue is our concern to determine the relationship between seasonality and dengue outbreak. Seasonality of dengue can be further classified into two components: climate-driven seasonality and non-climate driven seasonality. According to Sippy et al. (2019), for climate-driven seasonality, it can be related with relative humidity, rainfall and also surrounding temperature. In terms of non-climate driven seasonality, factors such as the effect of school schedules and public holidays pattern, may also contribute to dengue seasonality. Seasons drive human behavior because people may be more or less likely to spend time crowded indoors or spread outdoors, which affects the exposure rates. This can be the result of weather conditions or a result of seasonal holidays, which affect school and work schedules (Sippy et al., 2019).

1.3 Objectives

The objectives of the project are listed as below:

- To determine the dengue's outbreak season.
- To formulate the dengue outbreak incorporating the seasonality model.

1.4 Methodology

The methodology of mathematical modelling can be summarized into several steps.

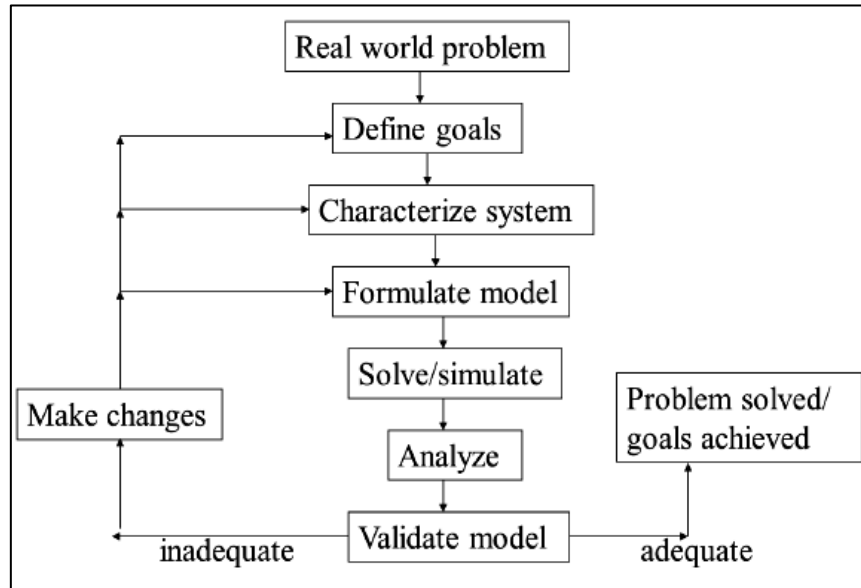


Figure 1.1: Summary on Methodology of Modelling Process

Figure 1.1 above shows the summary on methodology of modelling process. Methodology of modelling process involves several steps which consists of identifying the real world's problem, defining the goal related to the problem, characterization of system, model formulation, model simulation, analysis of the model and also model validation.

In this project, the problem statement is the seasonality of dengue. As the number of dengue cases increases, people are concerned whether seasonality-based modelling can pave the way for an early warning system. Thus, the goal on modelling of dengue seasonality is to identify the seasonality of dengue outbreak. After defining the goal, a list of assumptions related to the model will be listed out by identifying all the relevant data. The peak of dengue outbreak can be driven by climatic factors such as temperature, rainfall, humidity and others.

Next, model formulation is important to create a mathematical representation of a conceptual model. For model formulation, equation will be formulated to satisfy all the conditions and taking all climatic parameters into consideration. After model formulation, model simulation is required to determine the methods to solve the model, what calculations to make and also what kind of results to be obtained.

The methodology is then followed by the analysis of the model. In analysis part, data of visualization will be plotted to show the result the comparison of data. The final step on the modelling methodology is to validate the model. For model validation, the output or results of the formulated model is compared with the actual dengue cases occurred in Malaysia to determine whether the output of the formulated model is acceptable.

1.5 Scope

The field of this study regarding analysis of dengue seasonality is mainly focus in Malaysia. The scope of this project includes formulating and simulating a model. It also involves analyzing the available data between year 2010 to year 2017 on the number of dengue cases occurred in Malaysia. Dengue data is collected through Malaysian Open Data Portal (data.gov.my) and the official portal of Department of Statistics Malaysia (<https://www.dosm.gov.my/v1/>). Weather data is collected from Tutiempo. Tutiempo is a platform of world weather and local weather forecast. It is able to provide climate information for every country in the world with historical data. In this project, the relationship between dengue and seasonality is the main focus of study.

1.6 Significance of Project

The purpose of the project is for research and study on the relationship between dengue and seasonality. For instance, there is a study on the effect of rainfall for the dynamical transmission model of the dengue disease in Thailand. The result from the study showed that the most important factor for dengue virus transmission is the amount of rainfall as this

condition is most conducive for the development of Aedes mosquitoes. The relationship between average monthly dengue reported cases and average monthly amount of rainfall during 2003-2015 in Thailand metropolitan area is presented.

The significance of this project is just similar to the study as mentioned above. All of these studies are useful for the advanced development of seasonality-based modelling that could do the groundwork for an early warning system to minimize the loss. Different effect of seasonal variation is taken into account in this study. Besides, the impact of climate variability on dengue transmission in Malaysia such as temperature, total rainfall and average relative humidity are investigated. Hence, it is important to determine the dengue's outbreak season and also formulate the dengue outbreak incorporating the seasonality model. The outcome of this study will probably improve the effectiveness of dengue surveillance programs as well as controlling dengue epidemics.

1.7 Project Schedule

A project schedule has been created by using Microsoft Project 2019 to complete the Gantt chart of the tasks delivered. Start dates and finish dates are set to each task respectively which is from 16 September 2019 for Final Year Project 1 (FYP 1) and until 1 May 2020 as the submission date for complete Final Year Project 2 (FYP 2) report. Figure 1.2 below shows the Gantt chart of FYP 1 and FYP 2.

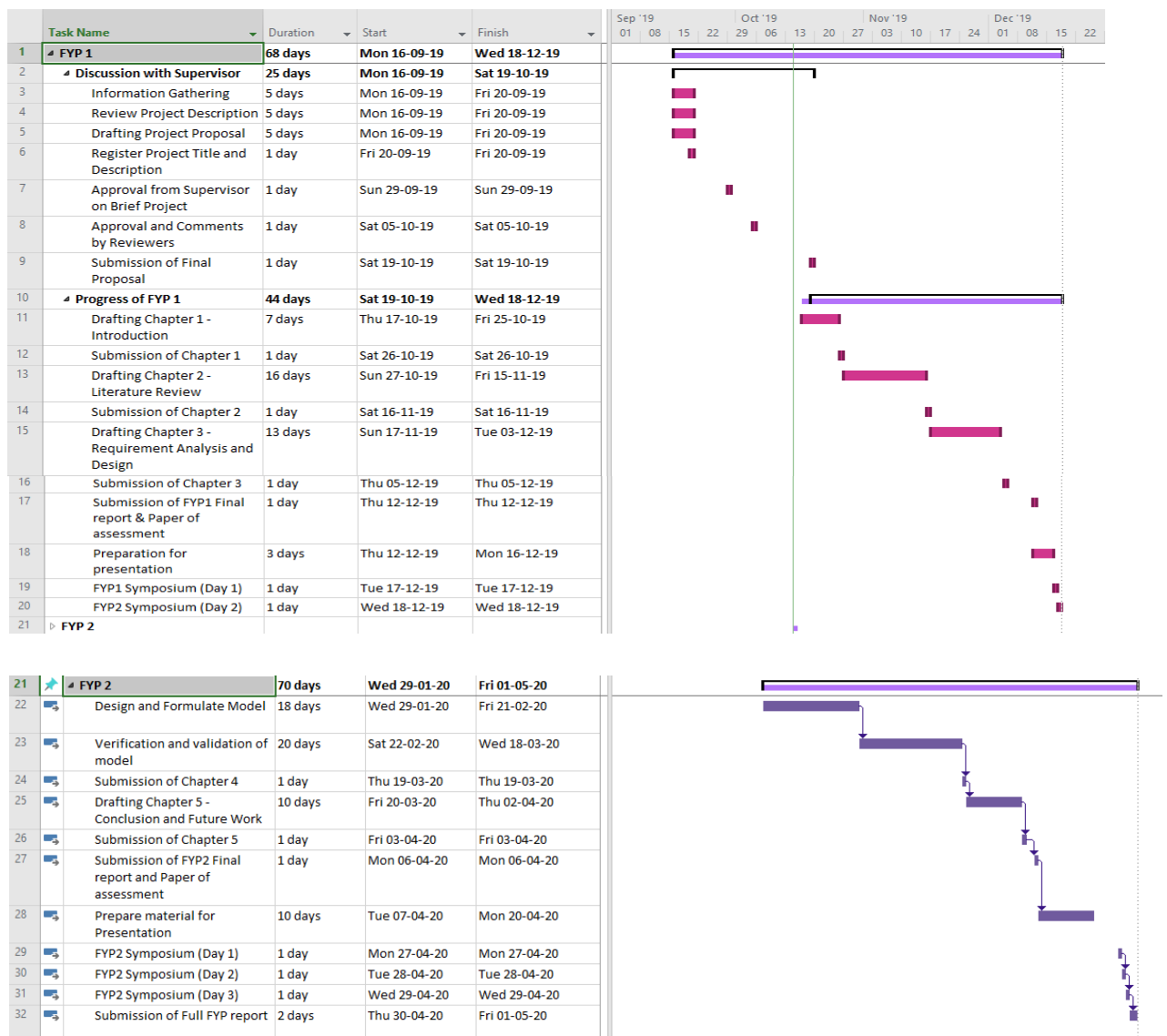


Figure 1.2: Gantt chart of FYP 1 and FYP 2

1.8 Report Organization

This report is organized as follows:

- Chapter 1 introduces the project by providing important details: overview, problem statement, objectives, methodology, scope, significance of project, project schedule, and project outcome.

- Chapter 2 includes the literature review which discuss some background knowledge for better understanding of the project.
- Chapter 3 discusses the methodology of achieving this project.

1.9 Project Outcome

Throughout this project, the outbreak season of dengue is determined. Besides, a model regarded to the dengue seasonality is formulated. The formulated model can be applied to predict the seasonality of dengue.

1.10 Summary

This chapter discusses about the introduction of the research. The introduction part also includes problem statement, objectives, methodology, scope, significance of project, project schedule, report organization and project expected outcome. The research is discussing about the relationship between dengue cases following epidemiological week and the climatic factor.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter begins with the review study of the relationship between the climate variables and the dengue. The background knowledge and review on the relationship between climate variables and the dengue is summarized in this chapter.

The section will be followed by three different approaches of dengue seasonality studies which includes machine learning, Autoregressive Integrated Moving Average (ARIMA), and compartmental model. Next, the section will discuss about the comparison between different approaches whereby different parameters are taken into consideration. Examples of existing models are discussed and compared to determine the best similarity with the project. Lastly, the data collection and the method used to carry out this project will be discussing. By discussing the data and method, the objectives of the research can be achieved.

2.2 Approaches of Dengue Seasonality Studies

Typical studies of dengue seasonality modelling aim to study the relationship between seasonal factors which relates to dengue outbreak. Researchers are interested to identify the seasonal pattern that are highly correlated to the outbreak of dengue. In term of seasonality, studies have been done on climate variables to examine its influences on dengue outbreak. For examples of climatic factors include temperature, rainfall and humidity. To date, there is a broad array of methods on dengue seasonality studies. For example, autoregressive integrated moving average (ARIMA) models, seasonal ARIMA, compartmental model (i.e. SIR model),

machine learning approach and others. These approaches are among of the most popular and further research can be made on them for better improvement.

2.2.1 Machine Learning

There are some dengue studies which have used scientific methods to determine the factor of seasonality related to dengue outbreak. Machine learning approach is one of the scientific methods. It uses computational methods to learn information directly from data without relying on a predetermined equation as a model. The performance of algorithms improve as the number of learning samples increases.

Machine learning uses two types of techniques as shown in Figure 2.1. The first technique is supervised learning, which trains a model on known input and output data so that it can predict the future outputs. The second technique is unsupervised learning, which finds hidden patterns or intrinsic structures in the input data.

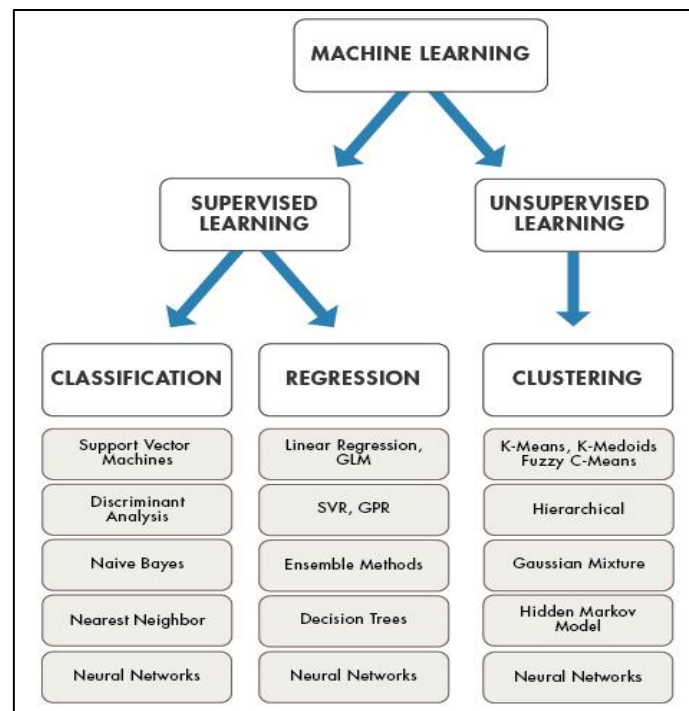


Figure 2.1: Types of machine learning (MathWorks)

New machine learning algorithms has been used by Stolerma et al. (2019) to forecast dengue fever in Brazil which relates to an assessment of climate conditions. The time series data for climate variables from seven Brazilian state capitals and their connection to dengue epidemic is analyzed in the study. All epidemiological data were taken from the publicly available datasets of the Brazilian Notifiable Diseases Information System. The datasets include the total number of dengue cases per year which is from year 2002 to year 2017 for all Brazilian state capitals.

Stolerma et al. (2019) analyzed that an epidemic year is when the incidence of dengue occurred at the specific city is above 100 cases as per 100,000 inhabitants in the period of January to December within a year, or otherwise classified as a non-epidemic year. The climate data utilized in this study was obtained from the National Institute of Meteorology (INMET) and included time series for the average temperature (in Celsius) and precipitation (in millimeters) for the seven state capitals which are Aracaju, Belo Horizonte, Manaus, Recife, Salvador, Sao Luis and Rio de Janeiro.

Support vector machine (SVM) was used by the researchers as a machine learning approach to forecast the dengue fever in Brazil. SVM is classified by machine learning as an example of supervised learning. The predictions for SVM made by Stomerlan et al. (2019) was based on a given set of predictors by looking for the relationship between independent variables and dependent variables. Furthermore, estimation is made on the correlation of climate conditions in different epochs preceding epidemic periods using a data-driven methodology based on machine-learning algorithms for clustering and classification. The estimation was then applied to climate variables which are key to the life cycle of mosquitoes. Seven state capitals that displayed at least 3 epidemic years and 3 non-epidemic years in the period of 2002 to 2012 are restricted in this paper. The SVM score for measuring discrepancies between