

UNIVERSITI TEKNOLOGI MARA

**DENGUE OUTBREAK PREDICTION
BASED ON METEOROLOGICAL
DATA USING MACHINE LEARNING
TECHNIQUES IN KOTA BHARU**

NOR FARISHA BINTI MUHAMAD KRISHNAN

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science
(Computer Science)

Faculty of Computer & Mathematical Sciences

October 2022

ABSTRACT

Dengue fever is a mosquito-borne infection that causes a high temperature, rashes, severe headache, muscle and joint discomfort, pain behind the eyes, and, in rare cases, bleeding. Rainfall, humidity, temperature, precipitation, floods, human movement, population, and the environment are only a few of the elements that induce dengue fever, including climatic and non-meteorological elements. This study used two different datasets that are dengue data and meteorological data that aims to identify the significant meteorological variables then develop a machine learning model to predict the dengue outbreak and proposed the machine learning. Dengue outbreak can be defined as 2 or more number of reported dengue cases in 7 days in certain regions. Random Forest feature selection is used to identify the significant meteorological attributes. It showed that maximum temperature, minimum temperature, average humidity and rainfall are significant for predicting dengue outbreaks. For modelling, Artificial Neural Network (ANN) and Decision Tree (DT) model were used to predict the dengue outbreak. Both models undergo parameter tuning to optimize the model. For ANN the different number of hidden nodes and decay were used to improve the model while for DT, maximum depth and complexity parameter were varying to improve the model. Both models, ANN and DT are evaluated based on accuracy, sensitivity and specificity showing that ANN (Accuracy = 69.05%, Sensitivity = 98.84%, Specificity = 3.80%), performed better than DT (Accuracy = 67.46%, Sensitivity = 97.11%, Specificity = 2.53%). The government and Vector Borne Disease Control (VBDC) may have preventive measures to handle the dengue outbreak as the meteorological parameters affect the dengue outbreak.

ACKNOWLEDGEMENT

Firstly, I am so grateful and wish to thank God for allowing me to embark on my Master and for completing this challenging journey successfully. I am so thankful and appreciation goes to my supervisor Dr Zuriani Ahmad Zukarnain and my co-supervisor Dr Azlin Ahmad and Pn. Marhainis Jamaludin. Thank you for your support, patience and ideas in assisting me with this study. Also, thank you to Dr Noorihan Abdul Rahman for your guidance with my study.

I would also like to express my gratitude to the staff of the Meteorological Department, especially En. Rosli Zakaria for providing the knowledge and assistance about the Meteorological data. Besides, I am so thankful to the Dr Syaarani Yasin and staff of Vector-borne Disease Control that assist me with dengue data. Special thanks to my friends for helping and guiding me with this study.

Finally, I would like to dedicate this thesis to my beloved mother and the loving memory of my very dear late father for the vision and determination to educate me. I am also thankful to my brothers Muhamad Faris and Muhamad Farid for giving me the support and motivating me to complete this thesis.

TABLE OF CONTENT

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENT	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER ONE INTRODUCTION	1
1.1 Research Background	1
1.2 Problem Statement	3
1.3 Research Objectives	4
1.4 Research Questions	4
1.5 Scope and Limitation of Study	4
1.6 Significance of Study	5
1.7 Thesis Outline	5
1.8 Summary	6
CHAPTER TWO LITERATURE REVIEW	7
2.1 Introduction	7
2.2 Dengue Fever Overview	7
2.3 Meteorological Elements Affect Dengue Outbreak	9
2.4 Dengue Outbreak Prediction using Machine Learning	14
2.5 Method of Analysis	22

2.6	Data Pre-Processing	23
2.7	Feature Selection using Random Forest	25
2.8	Artificial Neural Network (ANN)	29
2.9	Decision Tree	35
2.10	Performance Evaluation	40
2.11	Summary	42
CHAPTER THREE METHODOLOGY		43
3.1	Introduction	43
3.2	The flow of Research Methodology	43
3.3	Preliminary Study	45
3.4	Data Acquisition	45
3.5	Data Pre-processing	47
3.5.1	Data Cleaning	48
3.5.2	Merging Dengue Data and Meteorological Data	49
3.5.3	Handling Missing Values	51
3.5.4	Handling Outliers	52
3.5.5	Generate Target Variable	55
3.5.6	Data Transformation	56
3.6	Feature Selection using Boruta in Random Forest	57
3.7	Model Development for Dengue Outbreak Prediction	59
3.7.1	Splitting a Dataset into Train and Test Sets	59
3.7.2	Developing Artificial Neural Networks (ANN) Model	60
3.7.3	Developing Decision Trees (DT) Model	63
3.7.4	Parameter Tuning	65
3.8	Performance Evaluation	68
3.9	Summary	70
CHAPTER FOUR RESULT AND DISCUSSION		71