

UNIVERSITI TEKNOLOGI MARA

**PREDICTION OF PARTICULATE MATTER (PM)
2.5 FOR SPATIO-TEMPORAL DATASET USING TIME
SERIES MODELS**

ANIS MUNIRAH BINTI RIZUAN

BACHELOR OF SCIENCE (Hons.)

MANAGEMENT MATHEMATICS

JULY 2022

Universiti Teknologi MARA

**Prediction of Particulate Matter
(PM) 2.5 for Spatio-Temporal Dataset
Using Time Series Models**

Anis Munirah Binti Rizuan

**Report submitted in fulfillment of the requirements
for Bachelor of Science (Hons) Management
Mathematics Faculty of Computer and
Mathematical Sciences**

July 2022

SUPERVISOR’S APPROVAL

**PREDICTION OF PARTICULATE MATTER (PM) 2.5 FOR
SPATIO-TEMPORAL DATASET USING TIME SERIES
MODELS**

By

ANIS MUNIRAH BINTI RIZUAN

2019422992

This report was prepared under the direction of supervisor, Azlan Bin Abdul Aziz. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfillment of the requirements for the degree of Bachelor of Science (Hons.) Management Mathematics.

Approved by:

.....

Azlan Bin Abdul Aziz

Supervisor

JULY 15, 2022

ABSTRACT

Monitoring air pollution levels, particularly Particulate Matter 2.5 (PM2.5), is now crucial for the environment. The PM2.5 concentration needs to be evaluated to implement haze prevention measures since it impacts human health and the economy. This study focussed on investigating the pattern of PM2.5 at four stations; Shah Alam, Selangor (CA20B), Klang, Selangor (CA21B), Sri Aman, Sarawak (CA63Q), and Minden, Pulau Pinang (CA08P). Data from the Department of Environment (DOE) Malaysia have been obtained from 2018 to 2020 with 1096 observations. This study aims to determine the best "win" model and produce forecast values with high percentage accuracy by using Time-series Cross-Validation. Five models and four error measures have been implemented in this study. There are Naïve model, Mean Model, Single Exponential Smoothing Technique, Holt's method, and Box-Jenkins model. While the error measures used are Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scale Error (MASE). To execute these models, RStudio version 4.0.5 is based on R programming language 4.2.0. The results show that the best "win" model for Shah Alam, Klang, Sri Aman and Minden is Naïve model, Single Exponential Smoothing Technique, Holt's Method and ARIMA(3,1,1), respectively. The results of the study also found that the forecast value for the four locations studied recorded a high percentage of forecast accuracy such as Shah Alam (84-98 forecast accuracy), Klang (92-98 forecast accuracy), Sri Aman, Sarawak and Minden, Penang recorded a percentage of forecast accuracy between 94 - 98 percent. The finding of this study will improve Malaysians' control practice and awareness.

Keywords: Particulate Matter 2.5 (PM2.5), Univariate time series, Time-series Cross-Validation, Forecast Optimization.

TABLE OF CONTENTS

SUPERVISOR’S APPROVAL	ii
STUDENT’S DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1	1
INTRODUCTION	1
1.1 Background of Study	1
1.2 Problem Statement	3
1.3 Objective of The Study	4
1.4 Scope of Study	4
1.5 Significant of The Study	5
CHAPTER 2	6
LITERATURE REVIEW	6
2.1 Air Pollution	6
2.2 Particulate Matter 10 (PM10)	7
2.3 Particulate Matter 2.5 (PM2.5)	9
SUMMARY	12
CHAPTER 3	14
METHODOLOGY	14
3.1 Phase 1: Data Acquisition	15