

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

Anis Munirah Rizuan ¹, Azlan Abdul Aziz ²

¹*anismunirah03@gmail.com*

²*azlan172@uitm.edu.my*

ABSTRACT

Monitoring air pollution levels, particularly Particulate Matter 2.5 (PM_{2.5}), is now crucial for the environment. The PM_{2.5} 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 PM_{2.5} 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), Mean Absolute Error (MAE), Mean Absolute Scale Error (MASE), and Mean Absolute Percentage Error (MAPE). 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.

Keywords: Particulate Matter 2.5 (PM_{2.5}), Univariate time series, Time-series Cross-Validation, Forecast optimization