SIIC031

Three days ahead prediction of daily average concentration of PM$_{10}$ using regression tree approach

ᵛ *Firhad Dhiyafique Harun[1], Wan Nur Shaziayani Wan Mohd Rosly[2,*] and Prof Madya Dr Ahmad Zia Ul-Saufie Mohamad Japeri[2,3]*

[1] *Faculty of Chemical Engineering, Universiti Teknologi MARA, 13500 Pulau Pinang, Malaysia*

[2] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 13500 Pulau Pinang, Malaysia*

[3] *Hybrid Nanomaterials, Interfaces & Simulation (HYMFAST), Faculty of Chemical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500 Pulau Pinang, Malaysia*

*corresponding author: [2]shaziayani@uitm.edu.my

Abstract:

The air pollution in Malaysia are always fluctuated throughout the year. This is corresponding to the growth of industrial area and emission from vehicle. These episodes had indirectly significant impact on the air quality of Malaysia. Hence this study aims to study the trend and predict the 3 days ahead of daily concentration of PM$_{10}$ by using IBM SPSS and rapidMiner Studio respectively. The station is located at Jerantut, Pahang as it is significant to the centre of Peninsular Malaysia. As for the trend, air pollution hourly, monthly, and yearly monitoring records from 2004 until 2017 were used in analysing the statistical data analysis. There are eight parameters were selected in this study which known as PM$_{10}$, CO, SO$_2$, NO$_2$, O$_3$, relative humidity (RH), temperature (T) and wind speed (WS). The results obtained for the trend shows that it is higher reading of PM$_{10}$ in June until September every year due to the factors are usually coming from the neighbouring country and wind direction projectile to Malaysia. For the prediction, the results obtained for Root Mean Squared Error are 10.164, 13.853, and 13.281 respectively for day 1, day 2 and day 3. As for absolute error, the results obtained are 5.893, 8.268, and 9.052 respectively for day 1, day 2 and day 3. The results also indicated that dispersion of PM$_{10}$ in Malaysia were significantly affected by temperature, wind speed and relative humidity.

*Keywords***:**

PM$_{10}$, Regression Tree, Root Mean Squared Error, Absolute Error, rapidMiner Studio

Objectives:
- To determine the characteristic of PM$_{10}$ concentration by using SPSS Software.
- To predict PM$_{10}$ concentration using classification and regression tree by using RapidMiner Studio educational.

Methodology:

Data acquisition

Data of air quality monitoring in Jerantut, Pahang Malaysia was acquired from Department of Environment(DOE) Malaysia. The duration of data collection were from 2004 until 2017. The air pollutant parameters used in this work were PM_{10}, CO, SO,NO_2, O_3 and several meteorogical factors which known as wind speed, relative humidity and temperature.

Site Location

The location of sampling stations are chosen at air monitoring station ((N03° 58.238', E102° 20.863') which is at Jerantut, Pahang. This station is in fact a background station established by the Malaysian Department of Environment. This monitoring station is located at the Malaysian Meterological Deparment at Batu Embun, Jerantut, Pahang in the middle of the Malaysian Peninsular.
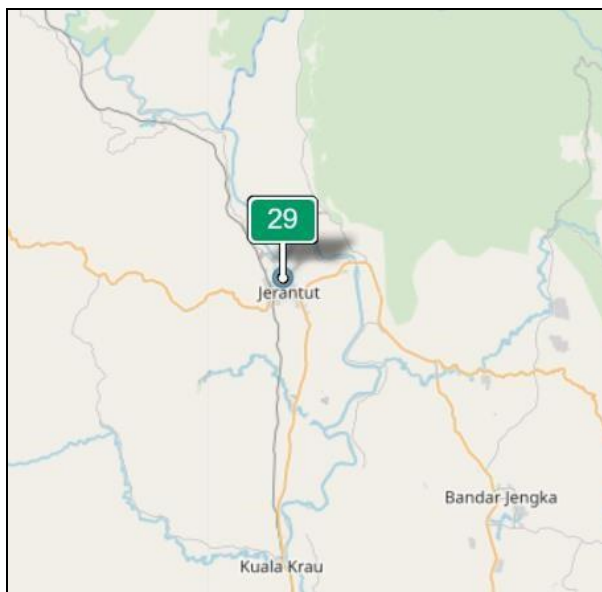

Figure 1: Air Monitoring Stations at Jerantut, Pahang

Statistical Analysis

The statistical analysis use a equation from SPSS and rapidMiner. The equation shows that the analysis as in the mathematics equation modelling. calculated computerized by the rapidMiner Studio when running the the rapidMiner Studio.

Results:

Prediction of PM_{10} concentration using classification and regression tree by using RapidMiner Studio educational.

Table 1: : Criteria and the results gain after performing regression tree modelling by using RapidMiner Studio

Criterion	Day- 1	Day-2	Day-3
Root mean squared error (RSME)	10.164 +/- 2.584	13.853 +/- 2.661	13.281 +/- 1.276

Absolute error	5.893 +/- 0.441	8.268 +/- 0.587	9.052 +/- 0.242
Relative error lenient	11.95% +/- 0.31%	16.21% +/- 0.20%	17.92% +/- 0.29%
Squared error	108.640 +/- 55.374	197.565 +/- 72.672	177.699 +/- 35.126
correlation	0.876 +/- 0.062	0.728 +/- 0.065	0.712 +/- 0.091

Conclusion:

 In conclusions, the objective to determine the characteristic of PM_{10} concentration by using SPSS Software is achievable. This is because the characteristic of PM_{10} can be determine by hourly, daily, and monthly by using SPSS Software. This is shown by getting the results of average concentration of PM_{10} in hourly data. This is because the weather that Malaysia's endured is raining and heat all year long which resulting in the reading of PM_{10} is within the range of +/- 1.00 throughout the day and night. Furthermore, the range of PM_{10} is almost the same as it still within the range of +/- 1.00. The value of PM_{10} concentration is fluctuate between the range of 38.