PREDICTION OF AIR QUALITY AT KLANG, SELANGOR USING MULTI-LAYER PERCEPTRON

BY

ANAS BIN AYUB
(2019361725)
MUHAMMAD ISMAIL BIN ELIAS
(2019338721)

UNIVERSITI TEKNOLOGI MARA

PREDICTION OF AIR QUALITY AT KLANG, SELANGOR USING MULTI-LAYER PERCEPTRON

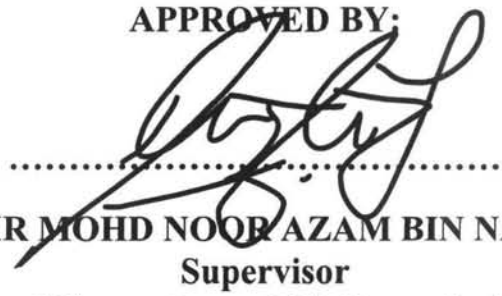
ANAS BIN AYUB
MUHAMMAD ISMAIL BIN ELIAS

Dissertation submitted in fulfilment
of the requirements for the degree of
Bachelor of Science (Hons) Statistics

Faculty of Computer and Mathematical Sciences

February 2021

APPROVED BY:



.....

(SIR MOHD NOOR AZAM BIN NAFI)
Supervisor
Faculty of Computer and Mathematical Sciences

ABSTRACT

This article aims to predict air quality by comparing various MLP functions. Artificial Neural Network as known as ANN architecture is the Multi-layer Perceptron (MLP) network is one of the most efficient predicting tools in many fields. Some of the elements of an MLP architecture is activation function. Picking the right activation on the MLP network plays a major role on the network performance. Hence, this study main objective is to compare 2 Multi-layer perceptron models with different activation function; Logistic function and Hyperbolic tangent function. This study also has identified the relationship between Gases and climatic variables with PM10 concentration. The dataset used in this study was gathered by Department of Environment Malaysia. The data is range from 2014 until 2018. Stepwise variable selection had been used to select variable to be used in the prediction models. The finding revealed that Carbon Monoxide has the highest correlation of 0.576 compared to other variables. There are only 4 variables used in prediction models: Carbon Monoxide, Humidity, Nitrogen Dioxide and Ozone. After compared the model using misclassification rate, average square error and ROC index, Multi-layer perceptron with Logistic function is the best model with Accuracy of 77.4% to predict the PM10 Concentration.

ACKNOWLEDGEMENT

Alhamdulillah, our utmost gratitude to Allah SWT for His guidance and in giving us strength, courage and persistence throughout our life, especially during difficult times in our life and with His consent we have the opportunity to complete this research

We would like to express our gratitude and appreciation to all those who have given us the opportunity to complete this report. A special thanks to our final year project supervisor, Mohd Noor Azam bin Nafi, whose help, stimulating suggestion and encouragement, helped us to coordinate our project especially in writing this report. We also want to express our sincere gratitude towards our Research Methodology lecturer, Dr. Wan Zakiyatussariroh binti Wan Husin for giving us the important information in conducting a research. Moreover, many thanks go to the head supervisor, Madam Nor Azima binti Ismail for handling the process of final year project proposal approval. We also would like to acknowledge the Department of Environment Malaysia (DOE) for accepting our request to use their dataset for our project.

We would like to thank our beloved family for giving us moral support especially during this pandemic times. Lastly, special thanks to our colleagues and friends for helping us directly and indirectly in completing our research. Alhamdulillah

TABLE OF CONTENTS

CHAPTER THREE: METHODOLOGY

CHAPTER FOUR: RESULT AND ANALYSIS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION LIST

No	ABBREVIATION	Name
1	PM_{10}	Particulate matter having aerodynamic diameter of below 10 micrometres
2	CO	Carbon monoxide
3	NO_2	Nitrogen dioxide
4	SO_2	Sulphur dioxide
5	O_3	Ozone
6	AT	Ambient temperature
7	RH	Relative humidity
8	WS	Wind speed
9	MLP	Multi-Layer Perceptron
10	ANN	Artificial Neural Network
11	API	Air Pollution Index
12	HRV	heart rate variability
13	DOE	Department of Environment of Malaysia
14	AQMS	Air Quality Monitoring Station

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND OF STUDY

In the past few decades, Malaysia has experienced tremendous industrial growth that has always been its goal to become a developed nation in the future. This phenomenon has been indeed a benefit to our economy. However, the actions led to emission of harmful gases and particulates in the air which leads to air pollution. Some air pollutants are poisonous. Inhaling them can increase the chance of health problems you will have. People with heart or lung disease, older adults, and children are at higher air pollution risks. Numerous studies have been done to investigate the association of air pollutant and heart rate variability (HRV). About 60–86% of these studies found that there are Negative associations across all HRV indices for periods of exposures to the pollutant ranging from 5-min to 5-days (Buteau & Goldberg, 2016).

Particulate matter is a tiny solid and liquid particle that floats in air. As the particulate matter floats or suspends in the air, it can be spread through air. Meteorological variables such as air temperature, relative humidity, wind speed, atmospheric pressure and amount of rainfall have had a significant impact on atmospheric PM10 concentration (Querol et al., 2004). PM10 is a particulate matter which has an effective aerodynamic diameter below 10 μm. The PM10 is one of the most dangerous pollutants; indeed, high levels of PM10 were correlated with increasing hospital admissions for lung and heart disease (Biancofiore et al., 2017). As exposure towards PM10 can affect human's health conditions, an essential step needs to be done. Prediction is one good way to know the upcoming reading of an item of interest. Therefore, prediction is an essential method for humanity to know the reading of upcoming PM10 concentration so that they can take appropriate steps to overcome this situation.

Air pollution status in Peninsular Malaysia is dominated by particulate matter, proven always having the highest Air Pollution Index (API) value compared to the other pollutants in most parts of the country (Abdullah et al., 2017). API is an index for reporting air quality on a daily basis. API used to measure how the local air quality affect health condition.

Thus, a few professors created and simulated several models to predict PM10 concentration in 1 day, 1 week and 2 weeks ahead (Dedovic et al., 2016). There were also some researchers created various prediction models using a Multi-Layer Perceptron with different input in order to measure accuracy of predicted PM10 concentration (Schornobay-Lui et al., 2019). Researchers at Malaysia also conduct research on developing the prediction models. There are models created to predict the PM10 concentration at Kuantan Pahang as it is monopolized by the tourism industry and petrochemical industry (Abdullah et al., 2020).

There are many methods that can be used to make a prediction. However, among all the methods, Artificial Neural Network (ANN) is one of the most efficient predicting tools in many fields. The group of statistical models is comprised of artificial neural networks (ANNs). It is worth noting that the use of neural networks has various possibilities. They are used not only in air protection but also in other environmental, economic, medicinal, industrial, etc. sciences. (Tadeusiewicz & Dobrowolski, 2004). ANN has two key benefits which is that it has the potential to function with limited information and by commenting on similar events, it can observe events and take decisions. One of the structures of ANN is the Multi-Layer Perceptron. One can obtain a Multi-Layer Perceptron (MLP) by applying a hidden layer to a simple perceptron. This architecture is typically trained using the Back Propagation of Errors (BP) algorithm or using any of its variants, so that in many cases the set generated by the MLP architecture plus BP learning is sometimes referred to as backpropagation of the neural network (García Nieto et al., 2012).

This study focuses on PM10 concentration as it is associated with air pollution. As this study focus on making prediction model to predict the level of PM10 concentration, it is beneficial to related authorities in making early necessary steps to prevent any unwanted disease regarding air pollution.

1.2 PROBLEM STATEMENT

There have been several newspapers this year posting articles regarding the harm of air pollution and the problems that people will face regarding air pollution. In the article "We're Unaware That Air Pollution Is A Silent Killer" posted by New Straits Times on 29th August 2019, it told us how air pollution can cause death and it is responsible for one out of nine deaths in Malaysia (Ahmad, 2019). As the government has put effort to solve this problem, the awareness among Malaysians of its harmfulness is still lacking. There was also a journal which found evidence that ambient PM2.5 concentrations measured in recent decades are associated with small but measurable increases in mortality from lung cancer (Turner et al., 2011).

Other than that, the Star Online also posted an article on 10th March 2020 regarding this issue. In the article, an author Jos Lelieveld from Max Planck Institute in Mainz, Germany stated that "Air pollution is a larger public health risk than tobacco smoking". Thomas Munzel from departments of chemistry and cardiology of the same institutes also said that there can exist a large scale of air pollution which they named it 'air pollution pandemic' (Zolkepli, 2020).

The statistical analysis showed the hourly trends (1-hour averaging time) of PM10, CO, O3 and NO2 in the Klang Valley were below the Malaysia Ambient Air Quality Guideline (MAAQG) standard. From table 1.1, Klang was the city with the highest number of days having the PM10 value > 100 $\mu g/m^3$; 14 days in 2007, 37 days in 2008, 39 days in 2009, and 41 days in 2011. The city recorded PM10 > 150 $\mu g/m^3$ for nearly half a month a year (12 days in 2009) (Latif et al., 2015). The reading of PM10 concentrations on Klang were undeniable as Klang is industrial and shipping sites.

Table 1.1: *Total Number of Days Between 2007 And 2011 When the Pm10 Standards Exceeded MAAQG*

Areas	Total Number of MAAQG Was Exceeded									
	2007 N=265		2008 N=265		2009 N=265		2010 N=265		2011 N=265	
	>100 µg/m³	>150 µg/m³	>100 µg/m³	>150 µg/m³	>100 µg/m³	>150 µg/m³	>100 µg/m³	>150 µg/m³	>100 µg/m³	>150 µg/m³
Klang	14	2	37	8	39	12	8	0	41	3
Petaling Jaya	0	0	0	0	13	0	0	0	12	0
Shah Alam	2	0	5	0	18	0	4	0	21	0
Kuala Selangor	8	0	4	0	39	10	0	0	21	0
Putrajaya	0	0	0	0	9	0	3	0	3	0
Cheras	0	0	0	0	21	1	0	0	8	0
Batu Muda	NA	NA	NA	NA	18	1	0	0	11	0
Banting	NA	NA	NA	NA	NA	NA	2	0	13	0

Source: (Latif et al., 2015)

According to "Air Quality Standards" by European Commission Environment, published on 31/12/2019, European Union standards and directives stated that maximum safe limit values for PM10 concentration in air is an average of 40 $\mu g/m^3$ annually and 50 $\mu g/m^3$ for an average of 24 hour. The daily limit should not be exceeded more than 35 times per calendar year. Thus, this makes Klang Valley one of the areas in Malaysia that have dangerous level of PM10 concentration.

As there already exists research in Malaysia to predict PM10 concentration using Multi-Layer Perceptron (Abdullah et al., 2020), the model used a data set acquired from Air Quality Monitoring Station at Kuantan. Thus, this model may only be suitable to predict PM10 concentration at Kuantan, Pahang only. Moreover, the data set period used from 2010 until 2014. The period of the data set raises a question whether the model can make an accurate prediction of PM10 concentration at Klang, Selangor.

Therefore, a thorough study needs to be done regarding this horrendous issue in Malaysia. Hence, this research has been conducted in the area of Klang Valley because the state has been involved in air pollution for past years.

1.3 RESEARCH OBJECTIVES

There are three objectives in this paper. These objectives must be stated to know whether the guidelines have been followed and the objectives have been successfully achieved. The objectives are as follows:

1. To identify the relationship between Gases (Carbon monoxide, Nitrogen dioxide, Sulphur dioxide, Ozone) and climatic variables (Ambient temperature, Relative humidity, Wind speed) towards PM10 concentration.
2. To determine the influencing variables to be used in Multi-Layer Perceptron to predict the PM10 concentration at Klang, Selangor.
3. To find the best model to predict PM10 concentration using a Multi-Layer Perceptron.

1.4 RESEARCH QUESTIONS

In this study, there are several questions that need to be answered so that this research can achieve its goal. The research questions are as follows:

1. What is the relationship between Gases (Carbon monoxide, Nitrogen dioxide, Sulphur dioxide, Ozone) and climatic (Ambient temperature, Relative humidity, Wind speed) towards PM10 concentration?

2. What are the suitable variables to be used in the Multi-Layer Perceptron to predict the PM10 concentration at Klang, Selangor?

3. What is the best model to predict PM10 concentration using a Multi-Layer Perceptron?

1.5 RESEARCH HYPOTHESIS

There are several initial hypotheses in this study. The hypothesis is based the preliminary assumption of this paper future findings when conducting this research. The hypothesises are:

1. The Gases and Climatic variables have strong relationship to PM10 concentration
2. Climatic variables (wind speed, humidity and temperature) and gases (Carbon Monoxide, Nitrogen Dioxide, Sulphur Dioxide) have a high impact in predicting PM10 concentration.
3. A model that uses hyperbolic tangent as activation function with the number of hidden nodes is the best model to predict PM10 concentration.

1.6 SIGNIFICANCE OF THE STUDY

As this issue keeps rising annually, a significant predictive model needs to be created. Thus, this research is very crucial as the objective of this research is indeed creating predictive modelling using Multi-Layer Perceptron. The finding of the research can be used by legitimate authorities like the Department of Environment Malaysia and Meteorological Department of Malaysia to make predictions of PM10 concentration so that a preliminary move can be made by the authority to curb the upcoming problem in the future. Preventing air pollutants is a must as it is harmful toward living things such as humans, animals and even plants.

1.7 SCOPE AND LIMITATION OF THE STUDY

This study aims to predict the air quality based on particulate matter sized 10 micrometres. Thus, this study does not take account of other particulate matter other than 10 micrometres. Moreover, this study uses a data set from the Department of Environment. This dataset originated from Klang, Selangor. Thus, the prediction model might only be suitable to predict the concentration of PM10 at Klang only. Thus, another prediction model needs to be calculated for other districts. This study has used variable that has only used variable that have been stated in previous study. Hence, other variables that might have connection to PM10 concertation had not be taken into

consideration. This map shows the general area of Klang Valley. The highlighted area is the district of Klang.

Figure 1.1: Map of Klang Valley. (source: Google Maps)

CHAPTER TWO
LITERATURE REVIEW

The purpose of this literature review in this report is to show the summary of the project being done and for essential information to understand and to be clear about the project. It explains the introduction of the project, the description of the problem regarding the current situation, a bit of information about the concentration of PM10 and how to predict it and finally the conclusion.

2.1 INTRODUCTION

Pollution is a phenomenon or event that frequently happens in Malaysia and any other developing countries. It can take forms of chemical substances or energy such as water pollution and noise pollution. Among all other pollutions, air pollution happens more frequently than any other pollution in Malaysia. As of today, air pollution has increased dramatically in developing countries such as Malaysia due to rapid population growth, the construction of many industrial buildings and the rapid urbanization (Talib et al., 2011). There are many sources of contamination from the air. In Malaysia, the three main sources of air pollution listed are mobile sources such as motor vehicle emissions, stationary sources such as power plant and factory emissions, and open burning (Afroz et al., 2003).

Air pollutants bring a lot of negative effects on a person's life. According to Buteau and Golberg (2016) as they have conducted a structured review of panel studies used to investigate associations between ambient air pollution and heart rate variability. They have found that 71–83% of studies of NO2 and 57–100% of studies of O3 stated that a person who was exposed to this pollutant from 5-min to 5-days will have a negative impact on their heart rate variability (HRV). Thus, they made a conclusion that air pollutants and HRV have negative associations.

2.2 PARTICULATE MATTER

Particulate matter is a very small solid and liquid airborne particle. PM10 is a particulate matter with a diameter of less than 10 μg/m3 that has a long-term demand for studies into the intricate nature of its accumulation of ambient air in the Malaysian Peninsula. air quality at Malaysia were dominated by particulate matter. This statement has been proven through the highest API value compared to other pollutants has been shown (Abdullah et al., 2017). Particulate matter is a small particle, both solid and liquid, that floats in air. It can be spread through air as the particulate matter floats or suspends in the air. Thus, meteorological variables such as air temperature, relative humidity, wind speed, atmospheric pressure and amount of rainfall have had a significant impact on atmospheric PM10 concentration (Querol et al., 2004). As such, any changes in these variables can affect the concentration of PM10 at certain places.

Because of its tiny size, particulate matter is harmful towards human body. Through the respiratory system, they can penetrate the human body, with larger particles retained in the upper respiratory tract, and those with smaller dimensions enter and damage directly into the lungs. The harmful effect of particulate matter pollution in young children and adults with chronic lung problems, pregnant women and neonates is most pronounced (Study et al., 2018). Long-term exposure to small particles of less than 10 micrometers can lead to a marked decrease in life expectancy. The decrease in life expectancy is mainly due to increased deaths from cardiopulmonary and lung cancers (Abdullah et al., 2017).

According to the Air Quality Guideline, the association between exposure to high concentrations of small particulates (PM10 and PM2.5) and increased mortality or morbidity is similar and measurable, both daily and over time. Conversely, as amounts of small and fine particulate matter decrease, associated mortality often decreases. PM10 therefore should not exceed the annual mean of 20 μg / m3 or the 24-hour mean of 50 μg / m3. When the average daytime reading of PM10 reaches 50 μg / m3 it means that the quality of the air is not healthy (World Health Organization, 2006).

2.3 PM10 PREDICTORS

In order to make an accurate prediction, suitable variables need to be used in the prediction model. Abdullah et al. (2020) have used meteorological variables and some gas component variables to make predictions of PM10 concentration. The meteorological variables are ambient temperature (°C), relative humidity (%) and wind speed (m/s). The gas components that have been used are carbon monoxide (CO, ppm), nitrogen dioxide (NO2, ppm) and sulphur dioxide (SO2, ppm). Other researchers have used these variables, but with additional variables to make the prediction. The total variables are particulate matter PM 2.5, nitrogen (II) oxide, nitrogen (IV) oxide, sulphur (IV) oxide, carbon (II) oxide, benzene, ozone (measured in air), lead, cadmium, nickel, arsenic, benzo(a)pyrene (Pawul & Śliwka, 2016).

In a previous study by Schornobay-Lui et al., (2019), they have used Temperature (°C), Relative humidity (%), Wind velocity (meters/second) and Precipitation (millimetres) to predict that particles. In the research, they conduct a comparison between models with different combinations of variables, and data sets. Because of that, some models have an additional variable which shows the value of PM10 concentration of the previous day. Not only that, local researchers also used these variables to make PM10 predictions. Past research has been done to develop and predict the next day, next two-day and next three-day PM10 concentration at Seberang Jaya, Malaysia also used particulate matter with an aerodynamic diameter less than 10 μm (PM10, μg m^{-3}) as its dependent variable. The independent variables that have been use are nitrogen dioxide (NO2, parts per million), wind speed (WS, km h-1), sulphur dioxide (SO2, ppm), Relative Humidity (RH %), Carbon Monoxide (CO, ppm) and ambient temperature (T, °C) (Ul-Saufie et al., 2015).

The use of meteorological data is important in making PM10 concentration prediction. The statement has been clarified by past researchers. In a previous study only four weather variables have been integrated into this model, namely wind speed, wind direction, temperature, and relative humidity. Wind speed and direction of wind, as shown in several studies such as Harrison et al., (1997), play a significant role in soil particle transport, dilution, and re-suspension. In addition, temperature and relative humidity have important effects on the concentration of PM10 (Branis and Vetvicka, 2010; Barmpadimos et al., 2011).

Hourly CO, SO2, NO, NO2, NO2 and PM10 concentration were among the parameters of the pollutants measured at the sampling site used in a research conducted by Sayegh et al. (2014). At the monitoring site of the Meteorology and the Environment (PME) Presidency, wind speed, wind direction, temperature, relative humidity, rainfall, and pressure are also monitored constantly.

According to Lešnik et al. (2019), they proposed a new machine learning method to allow a better prediction in terms of results accuracy and interpretability of the levels of PM10. For this reason, data have been collected systematically on an hourly basis from 2013 to 2016, including transport data (i.e., speed and category vehicles), weather data (i.e. temperature and humidity) and time stamps (i.e. the hour and day of the week).

McKendry (2002) uses the data from Chilliwack to build MLPs to estimate the maximum and average daily value of O3, PM10, and PM2,5. The inputs used on the models are hourly concentrations of NO, NO2, CO and O3. His study also Chilliwack weather variables (temperature and wind speed and direction), Annacis Island wind data, and Vancouver International Airport precipitation data.

2.4 MULTI-LAYER PERCEPTRON (MLP)

A conventional forecast approach is built on multivariate statistical analysis, but the application of neural artificial networks (ANN) has extended in recent years to make complex use of pollutant particulates (particularly PM10 and PM2.5). This is because its suseful way to fix systems like environmental contamination that do not have linear systems. The average daily PM concentrations in compliance with regulations as well as existing air Quality requirements were used in conjunction with weather variables (Biancofiore et al., 2017).

There are a variety of types of artificial neural networks that differ in structure and operating theory which known as multi-layer perceptron (MLP) or RBF network structure. The basic structure is made up of three separate neural layers neural artificial network (interconnected nodes). The first is the data-inserting input layer. The second is the hidden layer, which processes data to obtain the intermediate information needed

for calculating the final solution. The third layer is the output layer where it produces output (Pawul & Čliwka, 2016). To build the model, number of layers and the number of neurons must be determined. In the input layer, the number of neurons is equal to the number of independent variables to be used. To solve most classification problems, a hidden layer is necessary. The number of neurons in the hidden layer depends on the magnitude of the problem. While ANNs are used in complicated problems, in hidden layers more neurons are required. The number of neurons in the output layer corresponds to the number of predetermined groups or the output data number.

Multi-layer Perceptron (MLP) is a neural feed-forward network which can build highly complex nonlinear models, making it preferred in prediction of air pollution. The MLP starts when the desired input is fed into the network. This input parameter contains input signal, which is sent to the hidden layer from the input layer and to the output layer in the network from the hidden layer. The scaled input vector, which inserts neurons into the input stratum, is multiplied by weights (Abdullah et al., 2020). The multiplication of input and weight can be performed using activation function. There are several activation functions that can be used which are Linear, Hyperbolic Tangent, Logistic etc. the usefulness of this activation function depends on the data used.

ANNs consist of several processing components that resemble the neurons of the human brain and can perform smart tasks such as learning, generalising the information gained and patterns recognition. The literature provides excellent summaries with an atmospheric theme. In summary, the vast majority of physical science ANN's applications use the MLP model, consisting of an interconnected neuron system that allows nonlinear mapping (forwards) between a number of inputs and a number of outputs. MLP models were developed in the present application using the Statistical Software package (McKendry, 2002).

Multi Linear Regression is a statistical analysis that estimates the relationship between dependent and independent variables. The Multi-Linear Regression uses more than one independent variable. These statistical analyses are also being used by researchers to make prediction models regarding air quality. In a previous study by Biancofiore et al. (2017) The analyzed data corresponds to the Palacio de los Deportes (Sports Centre) station. This station has been chosen as one of the most populated areas of the city, where one of the motorways that enter the city is located. The research used

many methods to construct prediction models. The methods include multiple linear regression (MLR) models and a model with and without a recursive neural network architecture (NN). The best pattern was found for the two designs and PM10 concentration was compared for one, two and three days in advance. Then, PM2.5, which was measured more frequently, was predicted from the concentration of PM10.

The study compared MLP with MLR. The average daily PM10 concentration in a paper was predicted one, two and three days in advance by using three different models which were a recursive artificial neural network, a feed-forward neural system, and multiple linear regressions, showing that in all simulations, the recursive neural network model is more effective. Biancofiore 's study may conclude that the artificial neural network using Elman 's recurrent architecture (ANNE) performance exceeds both those of a neural network without recurrent architecture (ANNF) model and multiple linear regression (MLR) model. In all three forecasts, one, two and three days in advance, the neural network is showing improved performances for the MLR model with only meteorological inputs.

In addition, the ANN model was used as a valid example for estimating the concentration of hourly carbon dioxide in urban areas by studies in the Tabriz area. However, for the other model, the results were not satisfactory at all test stations. The ANN model presented can produce precise simulations of not just CO but other levels of air pollutants as well. This means that the Multi-Layer Perceptron is very suitable to predict particles in the air (Shakerkhatibi et al., 2015).

2.5 PREVIOUS STUDY USING ARTIFICIAL NEURAL NETWORK

Abdullah et al. (2020) acquired the data from the Department of Environment, Malaysia. The data was separated into two sections: training data set (Year 2010-Year 2012) and testing data set (Year 2013-Year 2014), respectively. Their study trained two MLP models with different activation functions in assessing the capability of the model for the prediction of air quality with the dependent variable of PM10 concentration. This study uses three sigmoid functions, for example, purelin (linear), log-sigmoid (logsig), and hyperbolic tangent sigmoid (tansig), which are commonly used as a part of MLP. The model was initialised at 5000 epochos and the model was run three times to achieve the mean output. The learning rate for the optimum configuration of the MLP network and improved model performance is set at 0.05.

Another study uses the meteorological data (maximum, minimum and average temperature, average wind speed, average preceding day temperature) and average daily concentrations of PM 10 of the preceding day as independent variables. There is one output of the neural network model which is PM 10 concentrations. All the experiments the data are randomly divided into three separate sub-sets: 75% for the training sub-set, 15% for the validation one and 15% for the testing one. The success measure used in these coefficients for correlation analysis (Pawul & Śliwka, 2016).

Schornobay-Lui et al., (2019) used the data occured in the city centre of São Carlos. Two sets called series 1 and series 2 were obtained from the data. Series 1 consists of daily collections from 1997 to 2006 while series 2 consists of daily collections from 2014 to 2015. The predictive models have been developed using meteorological data statistics. Two neural network architectures, namely the Multi-layer Perceptron (MLP) and Non-linear Exogenous Autoregressive (NARX), have been used to develop the models. The MLP were created 4 models with different combination uses of series data and with additional input (The day before PM10) or not. The first MLP with series 1 and no additional input. The second one is MLP with series 1 and additional input. The third one is MLP with series 2 and no additional input. The last one is MLP with series 2 and with additional input.

Another study conducted a comparison between two types of neural networks, multi-layer perceptron Feedforward Backpropagation (FFBP) and radial basis functions

(General Regression Neural Network (GRNN). The multi-layer perceptron that use Feedforward Backpropagation utilize PM10 as a dependent variable and Wind Speed, Relative Humidity, Temperature, Sulphur dioxide (SO2), Nitrogen Dioxide (NO2), past Particulate matter PM10 and Carbon Monoxide (CO) as independent variables.

According to Lešnik et al. (2019), PM10 concentrations measured by a laser aerosol spectrometer were automatically monitored with a wavelength of 660 nm in accordance with the EN 12341 Standard. The instrument was supplied with 31 size channels of particle counts and ready for measurements between 0.1 and 10.000 μg/m3. Due to maintenance work and calibration in the sensory system, the data were collected continuously from 01.01.2013 until 31.12.2016, except for short intervals, at a maximum of 2 to 3 weeks per year. The system's time resolution was an average of 1 minute. The sensory system measures traffic data by means of an induction loop simultaneously inside the road in addition to the concentration levels of PM10. The vehicle's presence and speed were measured with a 15-minute resolution and using a springboard inductance (Kwon and Parsekar, 2011) (i.e. by counting the number of vehicles and estimating their average speed). The mean number of vehicles per day was only over 8,000. Finally, the acquired data are combined in a time resolution of 1h with weather data.

2.6 CRITERIA TO EVALUATE ARTIFICIAL NEURAL NETWORK MODEL

According to Ling et al. (2007), Misclassification rates was defined as the number of misclassified events divided by the total number of events, were calculated for each testing data set. Misclassification rates are important indices for assessment of classification algorithms, as the classification aim is to lower the error rates of the test data and produce reliable predictions (Zhou et al., 2020). Misclassification rate defines how often the model is wrong in classifying. The formula for the misclassification rate defines the addition of FP (misclassify negative outcome as positive outcome) and FN (misclassify positive outcome as negative outcome) divided by all classifications:

$$\frac{FP + FN}{TOTAL}$$

According to (Yin & Tian, 2014) the area under the ROC curve (AUC) is a general measure of the accuracy of a biomarker / diagnostic test in the field of diagnostic studies. Another common index is the Youden index, defined as the correct overall classification rate minus one at the optimal cut off point. In the diagnostic field, the ROC curve and its associated statistics are very useful for assessing biomarker / diagnostic testing discrimination by continuous measurement. The region under the ROC (AUC) curve is a proven performance metric and calculates the likelihood of a random positive being ranked before a random negative, without setting the decision threshold. Often used to calculate the aggregate performance of the group on the basis that AUC averages total potential decision thresholds in some way. A plot of F1(t) (i.e., false positive rate at decision threshold t) on an x-axis against F0(t) (true positive rate at t) on the y-axis is defined as ROC (Swets et al., 2000; Fawcett, 2006 as mentioned on Flach et al . , 2011) with both amounts monotonously undeclining when t increases.

$$AUC = \int_0^1 F_0(s)dF_1(s) = \int_{-\infty}^{+\infty} F_0(s)f_1(s)ds$$

Samsuri, Marzuki and Al-Mahfoodh (2020) used several Indicators to assess the best fit models. The metrics used are the Root Mean Square Error and the Coefficient for Correlation. The precision measure shows that when the value is closer to 1 the most suitable model while when the error measure is closer to 0, the error measure indicates the better model. MATLAB R2015a was used to train and test a model.

The performance indicators that have been used are coefficient of determination (R2), Root Mean Square Error (RMSE), Index of Agreement (IA), Normalized Absolute Error (NAE) and Prediction Accuracy (PA) to evaluate the model (Ul-Saufie et al., 2015).

CHAPTER THREE
METHODOLOGY

3.1 SOURCE OF DATA

This research had used secondary data from the Department of Environment of Malaysia (DOE). The data from DOE had been requested through email attached with required documents. The required documents include DOE air quality data information application form and research confirmation letter from UiTM. The data requested from DOE were climatic variables (wind speed, humidity and temperature) and gases (Carbon Monoxide, Nitrogen Dioxide, Sulphur Dioxide). The data had been extracted from Air Quality Monitoring Station (AQMS) which originated at Klang that will consist of time series data with time interval of 24 hours and will be ranging from 2014 to 2018.

3.1.1 Type of Variable and Measurement

The data obtained consist of various independent variables and a dependent variable which is particulate matter having aerodynamic diameter of below 10 micrometres (PM10). The dependent variable (PM10 concentration) had been grouped into two categories according to its concentration which are:

Group A: PM10 with more than 50 $\mu g/m^3$

Group B: PM10 with less than 50 $\mu g/m^3$

The cut-off value is 50 $\mu g/m^3$ as it was stated by Air Quality Guideline that PM10 should therefore not exceed the annual mean of 40 $\mu g / m^3$ or the 24-hour mean of 50 $\mu g/m^3$ to maintain safe air quality. The detail of variables is shown in table 3.1.

Table 3.1

Characteristics of Variables

No.	Variables	Definition	Unit
Dependent Variable			
1	PM_{10}	Particulate matter having aerodynamic diameter of below 10 micrometres	$\mu g/m^3$
Independent Variables			
2	CO	Carbon monoxide	Parts per million (ppm)
3	NO_2	Nitrogen dioxide	Parts per million (ppm)
4	SO_2	Sulphur dioxide	Parts per million (ppm)
5	O_3	Ozone	Parts per million (ppm)
6	AT	Ambient temperature	Degree Celsius (^oC)
7	RH	Relative humidity	Percent (%)
8	WS	Wind speed	m/s

3.2 STUDY DESIGN

Using the data obtained, Statistical Analysis System (SAS) was used to construct multiple predictive models to predict the Particulate matter having an aerodynamic diameter of below 10 micrometres. Each of the models will use a different combination of activation function and a different number of hidden neurons. The dataset was allocated in the training set and the remaining set will be in the validation set. The training set had consisted of 70% of total observation and the remaining observation will be put in a validation set. After created all the models. Comparison model were conducted to identify which model is optimal for predicting PM10.

Klang, Selangor had been chosen as this paper's study area because it is one of the most populated areas in Malaysia. Moreover, Klang has many buildings and last year, Klang was also affected by the Trans-boundary haze incidents recorded in Central and West Kalimantan between 4–21 Sept 2019.

The theoretical framework displayed below shows the relationship of the variable that had been used. From the framework, the predictors used were Carbon monoxide, Nitrogen dioxide, Sulphur dioxide, Ozone, Ambient temperature, Relative humidity and Wind speed. All the predictors will be used to predict Particulate matter having an aerodynamic diameter of below 10 micrometres (PM10).

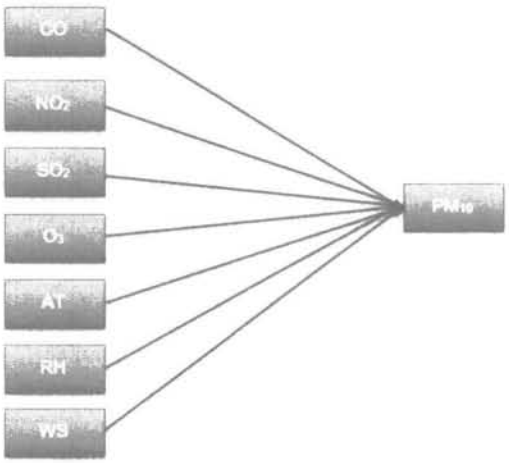


Figure 3.1: Theoretical Framework

3.3 STATISTICAL PACKAGE

In this paper, the software that had been used was SAS Enterprise Miner software which were to create and compare the predictive model and to process and analyse the data.

3.4 DATA PREPERATION AND DATA MANIPULATION

3.4.1 Identifying Outlier

Outlier can be defined as elements from a dataset that is far from other elements in the dataset. To put it in a simple term, they are abnormal values in a dataset. If detection of outlier is not done in a data preparation, it can be very problematic for many analyses because outlier can cause either miss significant findings or distort real results. Hence, there are many ways in detecting outlier in a dataset since there are no strict statistical rules for it. Detecting outliers depends on understanding the data during the collection process.

There are many ways in detecting outliers in a dataset including sorting the dataset and graphing the data such as histogram, boxplots and scatter plot. Figure 3.2 shows an example of outlier detection by using a histogram. There is an element from the dataset that is too far to the left of the graph.

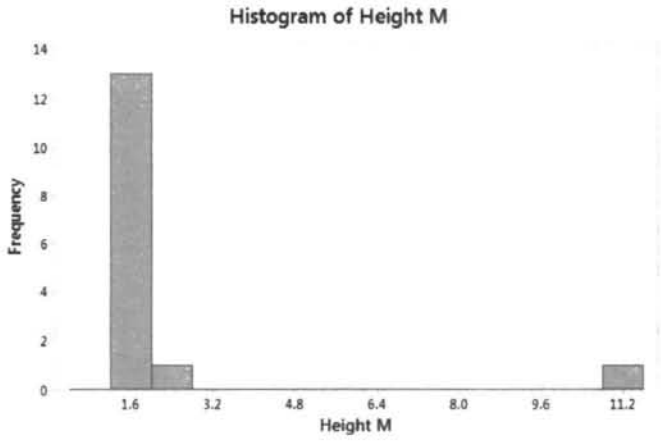


Figure 3.2: Example of Histogram in Detecting Outliers

3.4.2 Replacing Value

After detecting outliers, those extreme values need to get rid of. First, this study had replaced all the extreme value with missing value. Moreover, this step also gets rid of meaningless value in the variables by also replacing it with missing value. Meaningless value defined as value that did not have any purpose or significance in the dataset. This study used a node called 'Replacement node' in SAS Enterprise Miner to execute this step. The value that determine it outlier depend on each of the variables.

3.4.3 Impute Missing Value

After executing the replacement value step, all those missing values get replaced with average value correspond to where the missing value belong in each variable. However, Variables that consist of 50% of its observations as missing value will not be imputed instead the variables will not be used in this study. This study used node called 'Impute node' in SAS Enterprise Miner for this step. This step ensures to minimize the deletion of observations in dataset. Deleting many observations will reduce the size of dataset, hence it will impact the training of prediction models.

After executing the replacement value step, all those missing values get replaced with average value correspond to where the missing value belong in each variable. However, Variables that consist of 50% of its observations as missing value will not be imputed instead the variables will not be used in this study. This study used node called 'Impute node' in SAS Enterprise Miner for this step. This step ensures to minimize the deletion of observations in dataset. Deleting many observations will reduce the size of dataset, hence it will impact the training of prediction models.

3.4.4 Transforming Variable

The target variable for this study is PM10. Originally, the variable in dataset was in continuous variable form. To make binary prediction model, the target variable needs to be transformed into binary variable. Hence the original variable PM10 were grouped into 2 groups. The first group is PM10 with more than 50 $\mu g/m^3$ labelled as '1'and the second group is PM10 with less than 50 $\mu g/m^3$ labelled as '0'.

3.4.5 Dropping Variable

During the process of replacing and imputing values in the data set, SAS Enterprise Miner will create new variable of each process for each variable that was processed. Hence, the old raw variable needs to be dropped so that it cannot interfere with during the data analysis and the final result in the data analysis.

3.5 DATA ANALYSIS

3.5.1 Descriptive Analysis

In this study, descriptive analysis was used to describe and sum up the data. The data were either described using tabulation or a graphical approach. Since the data set consists only of quantitative data, no qualitative data analysis will be carried out. Hence, for quantitative analysis regarding mean, standard deviation, median (interquartile range) and histogram were applied.

3.5.2 Pearson Correlation Coefficient

This study used parametric Pearson correlation to find the linear relationship between dependent variable and independent variable. The correlation measures the strength relationship between two variables. The relationship variables that were used to test are between factors (CO, NO_2, O_3, ambient temperature, relative humidity, wind speed) and particulate matter having an aerodynamic diameter of below 10 micrometre (PM10). The range value of the correlation coefficient lies between -1.0 and +1.0. The strength relationship is summarized in the table 3.2:

Table 3.2
Strengths of relationship

Pearson Correlation Coefficient, r	Strengthness
>0.75	Perfect correlation
0.5<r<= 0.75	Medium correlation
0.25<r<=0.5	Fair correlation
<=0.25	Little/no correlation

3.5.3 Artificial Neural Network

According to Biancofiore et al. (2017), in all of the simulations in the paper, the indices suggest that the Artificial neural network using recurrent Elman architecture (ANNE) performances surpasses both those of neural network with no recurring architecture (ANNF) and of the multiple linear regression model (MLR) model. The neural network reveals better performance than MLR model in all three forecasts, 1, 2 and 3 days ahead, using only meteorological inputs.

A typical neural network structure is composed of an input layer, a hidden layer, and an output layer. Data enters the network through the input layer. The hidden layer consists of artificial neurons, each receiving multiple input from the input layer. Output layer is a layer that combines the artificial neurons summarizing results.

A neural network can have many hidden layers, but one hidden layer is generally enough. The wider the layer, the bigger the network's ability to recognize patterns. The neural network will be more able to memorize the pattern within the training set. This will result in overfitting. Hence, this study will utilize 5, 6, and 7 numbers of hidden nodes in hidden layers to identify which numbers of hidden nodes is suitable.

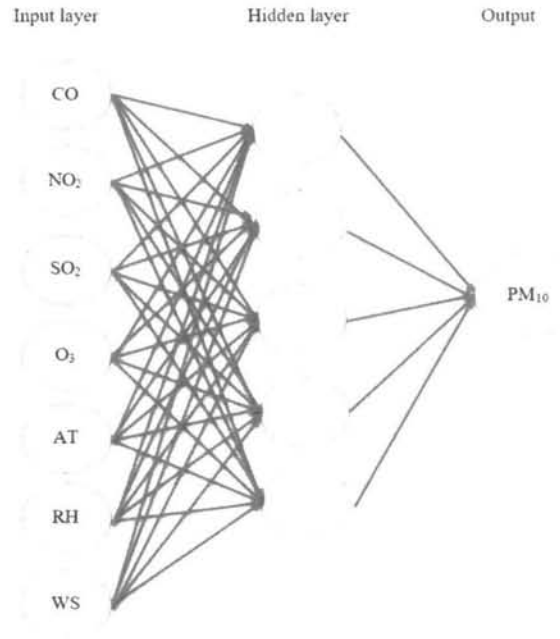


Figure 3.3: ANN Multi-Layer Perceptron Models

3.5.4 Variable Selection

In order to make accurate prediction model, some variable need to be excluded. Stepwise variable selection had been used to select the suitable variable. The Regression node had been used in SAS Enterprise Miner. The regression node had been linked to Neural Network node to enable the Neural Network node to use only selected variable.

3.5.5 Activation Function

The activation function has two part which are combination function and activation function. The combination function will emerge all input into a single value. In this study, the combination function in the hidden layer will use weighted sum where each input is multiplied by its weight and these products are added together. This study compared 2 activation functions which are Logistic function and Hyperbolic Tangent function.

3.5.5.1 Logistic Function

Logistic or sigmoid function can produce the value for the target variable between 0 and 1 (Sibi et al., 2013). The sigmoid activation function is most useful for training data. It is one of the most used activation functions. The logistic function is given as:

$$g(x) = \frac{1}{1+e^{-x}}$$

This function proven very usefull to use in neural networks trained by back-propagation algorithms. This is because it is easy to differentiate and can interestingly minimise the training computing capability (Karlik & Olgac, 2011). The term "sigmoid" means "S" and the logistic shape of the "sigmoid" means the interval (-∞, ∞) onto (0, 1) as shown in Figure 2.2.

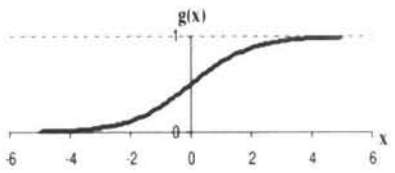


Figure 3.4: Graph of Logistic Function

3.5.5.2 Hyperbolic Tangent Function

Hyperbolic tangent function as an alternative for logistic function. According to Karlik & Olgac (2011), this function may be described as a relationship between the sinus functions of the bone hyperbolic and the cosines, or as the ratio of the half sum to the two exponential functions in points x and -x can be extended as follows: The following functions:

$$tan(x) = \frac{sinh(x)}{cosh(x)} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

Hyperbolic tangent function is similar to logistic function. It will show the range output between -1 and 1 as shown in the figure:

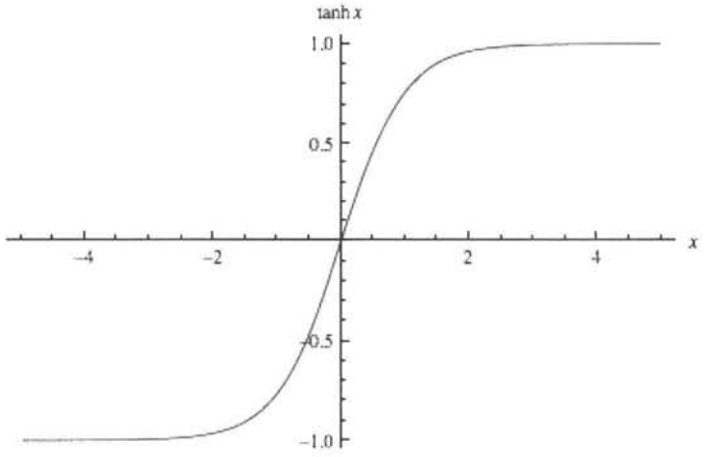


Figure 3.5: Hyperbolic Tangent Function for the real values of its argument x

The advantages of using the hyperbolic tangent function are that the negative inputs are highly negative and that in the tanh graph, the zero inputs are mapped to almost zero. The function is differentiable, and it is also commonly used in classification between two classes.

3.5.6 Model Structure in SAS Enterprise Miner

The structure had been developed in SAS Enterprise Miner. Replacement node had been used to replace any wrong value first. Then linked to the Data Partition node, this node separates the data into 2 partitions which are the training set and validation set. To handle missing value, the Impute node had been used to replace the missing value.

After that, 2 models had been created with different activation function. The two models had been compared based on 3 criteria, which are misclassification rate, average square error and ROC index. The flow of the analysis from SAS E-Miner canvas can be seen below:

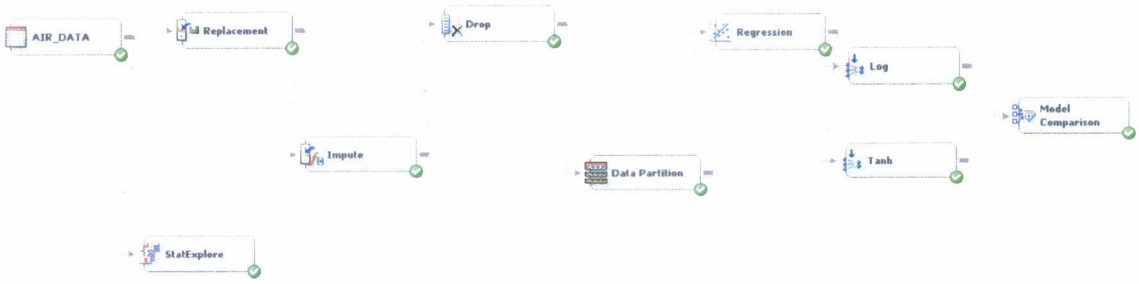


Figure 3.6: The Flow of The Analysis

3.5.7 Iteration training

An iteration is a term used in machine learning and shows how many times the parameters of the algorithm are updated. The exact meaning of that will depend on context. Neural network training will require a great many iterations because to rethink and adjust all weighting factors so that an accurate prediction can be done.

The maximum iteration that has been used in this paper is 8. For each activation function, the dataset will go through iteration 8 times to adjust all influential factors so that the final prediction of each function would be accurate.

3.5.8 Confusion Matrix

All the model created will be analysed according to its confusion matrix. Confusion matrix is a table that describe the performance of classification models. This study created confusion matrix for each of the model. Since this study used 2 set of data: training set and validation set, each model has to confusion matrix correspond to each set. From the confusion matrix, five measurement will be calculated which are: precision, misclassification rate, specificity, sensitivity, and accuracy. The formula for each of the measurement are as follow:

Table 3.3:

Formula of Performance Measurement

Measurement	Formula
Precision	$\frac{TP}{(TP+FP)}$
Misclassification rate	$\frac{FP+FN}{(TN+FP+FN+TP)}$
Specificity	$\frac{TN}{TN+FP}$
Sensitivity	$\frac{TP}{(TP+FN)}$
Accuracy	$\frac{TP+TN}{(TN+FP+FN+TP)}$

3.5.9 Model Comparison

In choosing the best model among all the models created, there are several criteria that must be fulfilled. The value of criteria would be compared between the training set and validation set. The criteria that has been used are misclassification rate, average square error and Receiver Operating Characteristics (ROC) Index. Each criterion would be examined in both training and validation set. The gap values of the criteria would be calculated and would be used to choose the best model.

In choosing the best model, it should have the lowest Misclassification rate and average square error. Meanwhile, the Receiver Operating Characteristics (ROC) Index have the highest value. This indicate the model will have minimal fault in misclassifying the outcome as well as making mistakes in predictions. The ROC Index should be high as it shows the performance of the model. Hence, higher ROC index indicates higher performance.

Table 3.4:

Example Table of Model Comparison

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion Valid: Average Squared Error	Train: Sum of Frequencies	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies	Valid: Maximum Absolute Error
Y	Neural Tree	Neural Tree	Neural N... Decision ...	Y Y	Projected... Projected...	6.843E-5 .0001326	17658 17658	0.05529 0.052587	1.220497 2.203752	6.912E-5 .0001248	0.008314 0.011171	17658 17658	17658 17658	13244 13244	0.043332 0.08293

3.6 SUMMARY OF DATA ANALYSIS TECHNIQUE

Table 3.5 shows the summary of data analysis techniques that were used in the study based on the objective of this study.

Table 3.5:

The Summary of Data Analysis Technique

OBJECTIVE	VARIABLE	DATA ANALYSIS TECHNIQUE
To identify the relationship between independent variables (Carbon monoxide, Nitrogen dioxide, Sulphur dioxide, Ozone, Ambient temperature, Relative humidity Wind speed) and dependent variable (Particulate matter having aerodynamic diameter of below 10 micrometres	• Carbon monoxide • Nitrogen dioxide • Sulphur dioxide • Ozone • Ambient temperature • Relative humidity • Wind speed • Particulate matter having aerodynamic diameter of below 10 micrometres	• Pearson Correlation of Coefficient
To determine the influential variables to be used in Multi-Layer Perceptron to predict the PM10 concentration at Klang, Selangor.	• Carbon monoxide • Nitrogen dioxide • Sulphur dioxide • Ozone • Ambient temperature • Relative humidity • Wind speed	• Stepwise variable selection

To find the best model to predict PM10 concentration using a Multi-Layer Perceptron.	• Carbon monoxide • Nitrogen dioxide • Sulphur dioxide • Ozone • Ambient temperature • Relative humidity • Wind speed • Particulate matter having aerodynamic diameter of below 10 micrometres	• Misclassification Rate • Average Square Error • ROC Index

CHAPTER FOUR
RESULT AND ANALYSIS

4.1 DESCRIPTIVE ANALYSIS

Table 4.1: Descriptive Table of all variables in the dataset

Variable	Mean	Median	Maximum	Missing	Non-Missing
CO	1.295	1.24	8.46	4	1741
Humidity	81.16	83.6	99	0	1745
NO2	0.033	0.034	0.083	23	1722
O3	0.044	0.043	0.115	25	1720
SO2	0.005	0.004	0.0624	0	1745
temperature	31.807	32.6	37.6	0	1745
Wind Speed	3.286	3.194	6.5	1199	546
PM10	67.474	58	466	1	1744

The table shows the standard characteristics of each variable which is median, minimum, maximum, mean value etc of the variable. The variable is CO concentration, humidity, NO2 concentration, O_3 concentration, the target variable which is PM10 concentration, SO_2 concentration, temperature, and wind speed. The median of each variable is 1.24, 83.6, 0.034, 0.043, 58,0.04, 32.6 and 3.194 respectively. Moreover, the mean of each value is 1.295, 81.16, 0.034, 0.044, 67.47, 0.0056, 31.807, 3.287 respectively. Next, the variable that has the most missing value is Wind Speed with 1199 missing values while CO, NO2, O_3 and PM10 concentration has 4,23,25,1 respectively.

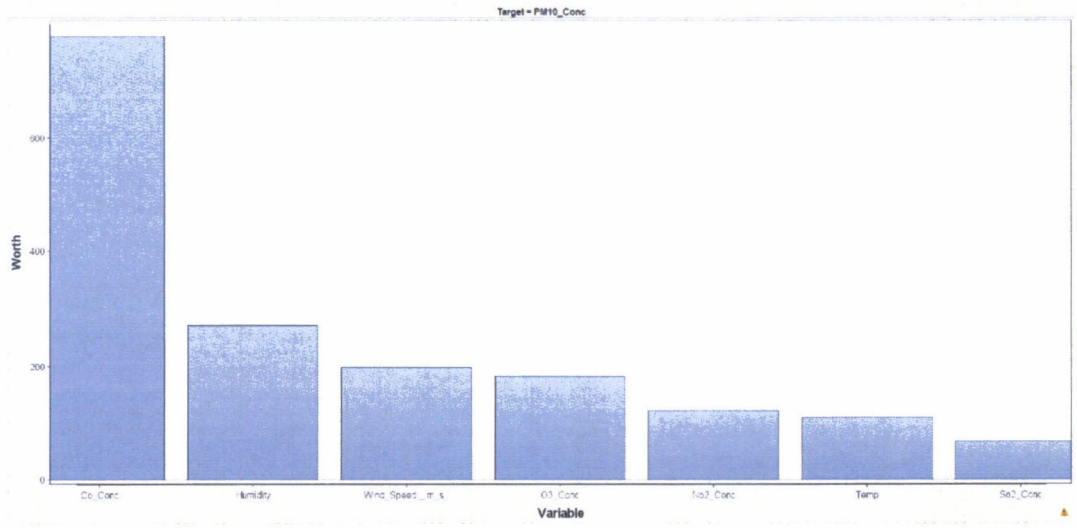


Figure 4.1: Worth Graph of All Input Variables

In Figure 4.1, the graph shows the worth value of each input variable that is worth to define the target value which is PM10 concentration. In the graph, the variable with the most worth value that defines the target value is CO concentration which is 774.25 and the input variable with the least worth value is SO2 with the value of 66.03.

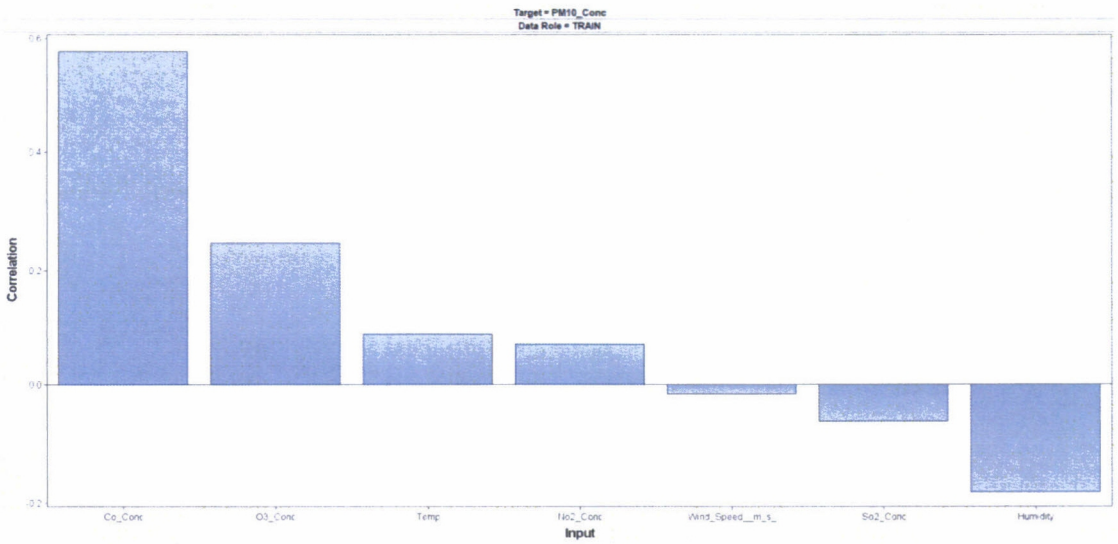


Figure 4.2: The Correlation Graph of All Input Variables

In Figure 4.2, the graph shows the correlation of each input variable to the target variable. This graph indicates the influences of each input variable towards PM10 concentration. The variable that has positive correlation to the target value is C0 concentration, 03 concentration, temperature, and N02 concentration while wind speed, SO2 concentration, and humidity has negative correlation to the target value. The variable with the most influence on the target variable is C0 concentration which is 0.576 correlation and the variable with the least influence is humidity with the correlation of -0.181.

4.2 DATA PROCESSING

In the descriptive analysis, there are some of the variable that has minimum value of 0, this does not make sense since the particle in air will have those even though if it's too little. As for Humidity level, it is impossible to have completely 0% relative humidity as water vapor is always present in the air, even if only trace amounts. As for the temperature, it is impossible to have 0 degree Celsius as Malaysia is in the Northern Hemisphere, which is the northern part of the Equator. The average temperature in almost all of Malaysia is between 21 ° C to 32 ° C. The graph below shows the missing value in each variable.

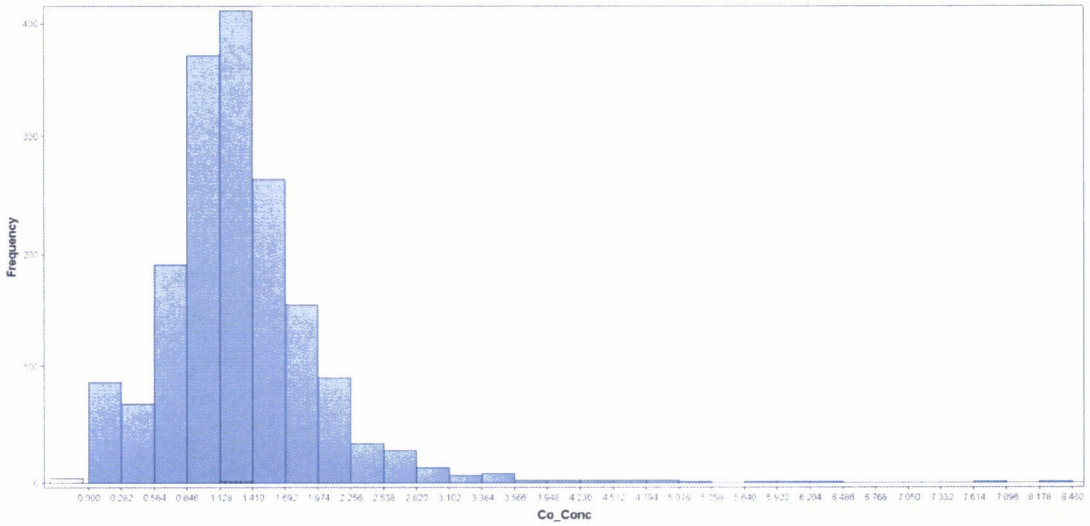


Figure 4.3: Histogram of Variable Carbon Monoxide

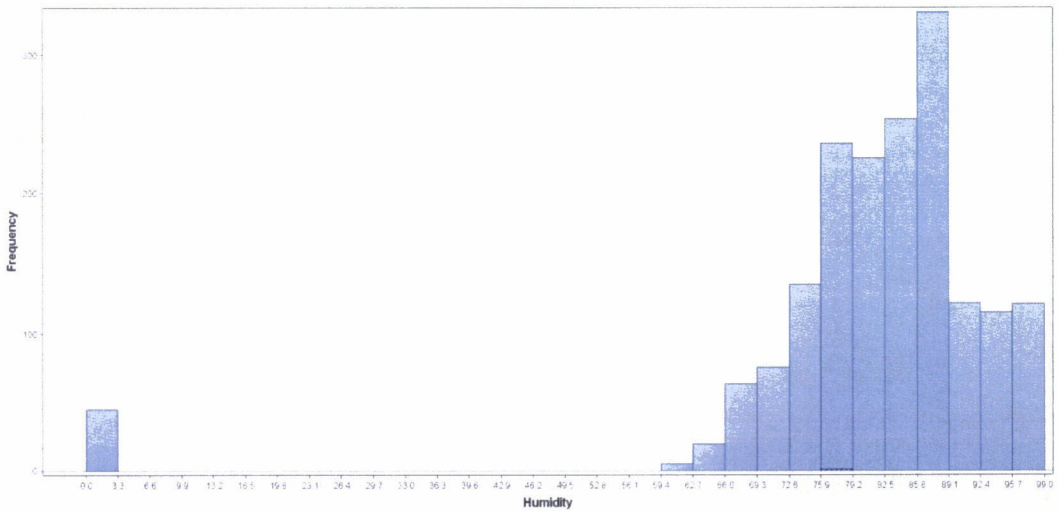


Figure 4.4: Histogram of Variable Humidity

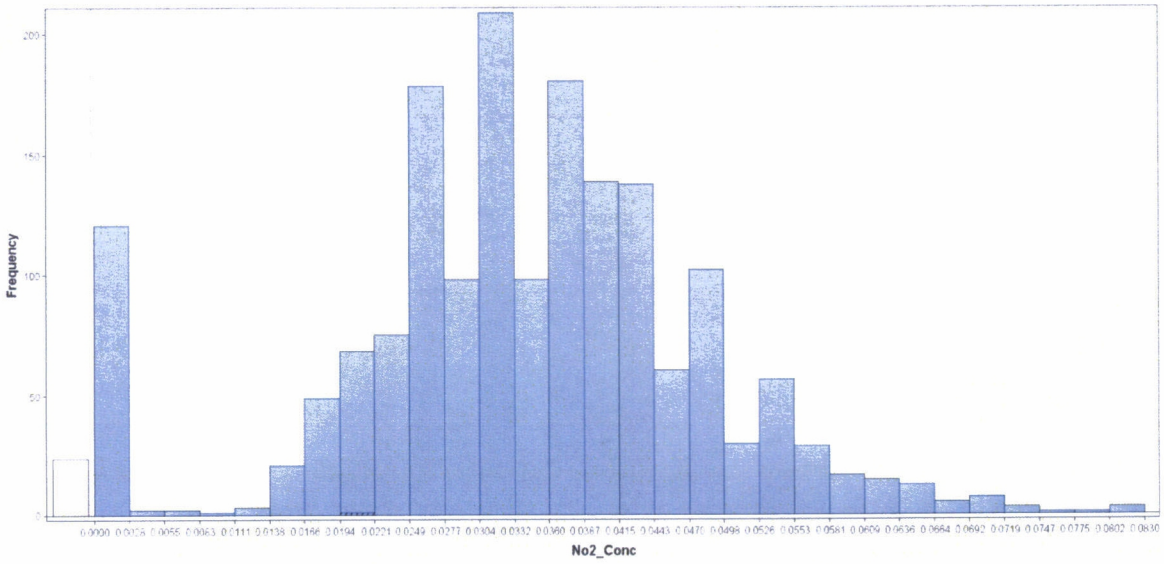


Figure 4.5: Histogram of Variable Nitrogen dioxide

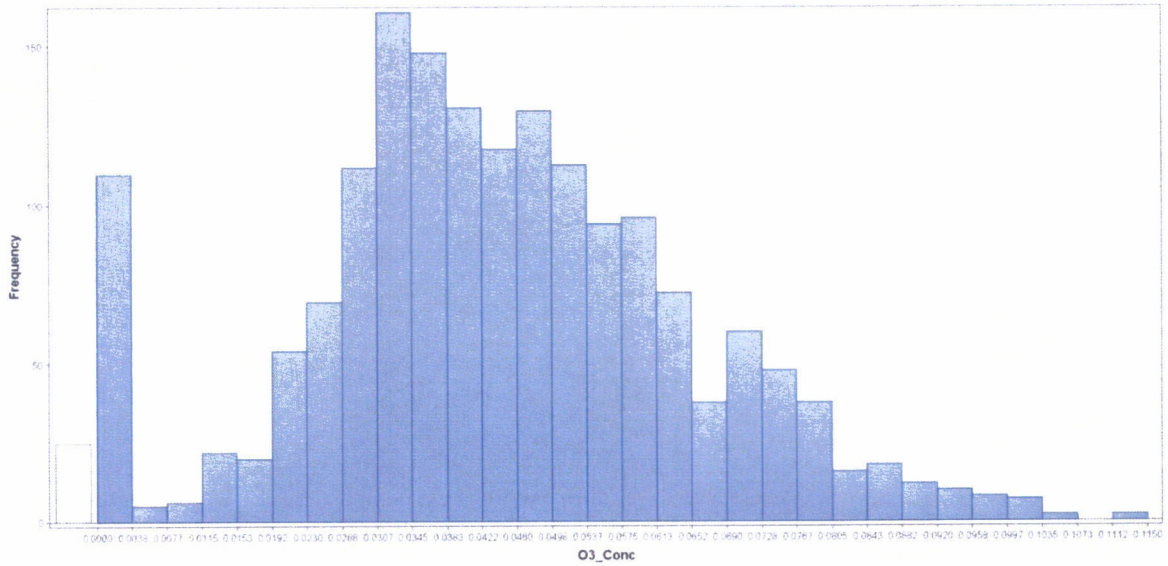


Figure 4.6: Histogram of Variable Ozone

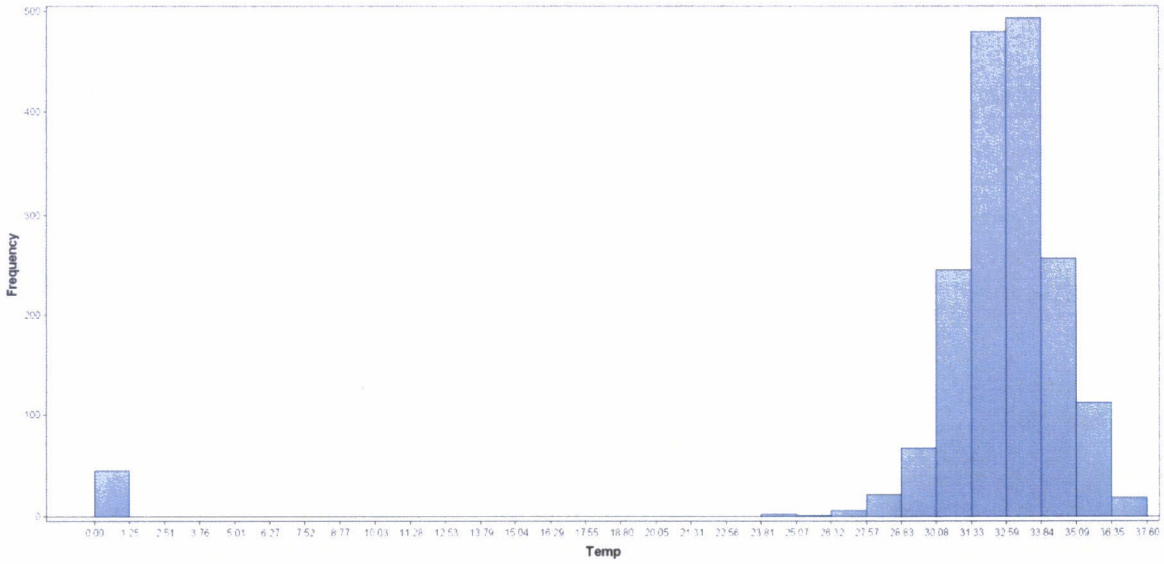


Figure 4.7: Histogram of Variable temperature

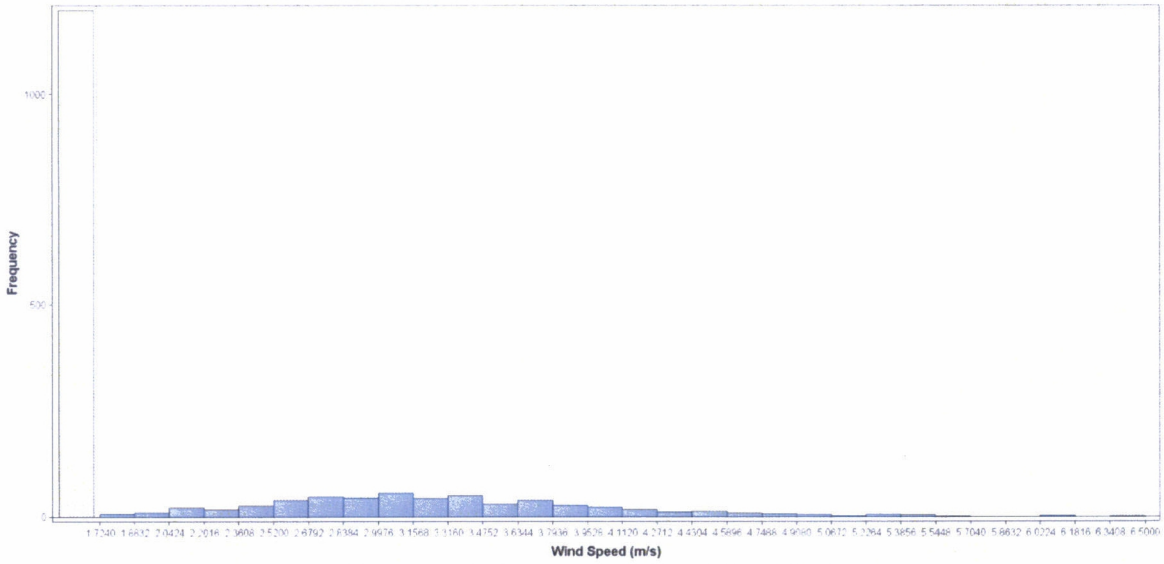


Figure 4.8: Histogram of Variable Wind speed

4.2.1 Replacement Value

To solve this, the unwanted value needs to be replaced to missing value. After that, the missing value will be imputed using mean value of each variables. The Replacement node in SAS Enterprise Miner were used to make value replacement. In the Replacement node, certain lower limit value has been set. If the original value falls below the lower limit value, the original value will be replaced with missing value. The CO, Humidity, NO2, O_3 and Temperature will have their own distinct lower limit value. The lower limit for CO, Humidity, NO2, O_3 and Temperature are 0.1, 5, 0.001, 0.001 and 2 respectively. After running, it shows that several variables that has meaningless 0 value has been replaced to missing value. All the variables are input variables. Carbon monoxide, Humidity, Nitrogen Dioxide, Ozone and Temperature now have additional missing value of 79, 44, 121, 108 and 44 values respectively.

Limits and Replacement Values for Interval Variables

Variable	Replace Variable	Lower limit	Lower Replacement Value	Upper Limit	Upper Replacement Value
Co_Conc	REP_Co_Conc	0.100	.	.	.
Humidity	REP_Humidity	5.000	.	.	.
No2_Conc	REP_No2_Conc	0.001	.	.	.
03_Conc	REP_03_Conc	0.001	.	.	.
Temp	REP_Temp	2.000	.	.	.

--
* Report Output
--

Replacement Counts

Obs	Variable	Label	Role	Train
1	Co_Conc	Co_Conc	INPUT	79
2	Humidity	Humidity	INPUT	44
3	No2_Conc	No2_Conc	INPUT	121
4	03_Conc	03_Conc	INPUT	108
5	Temp	Temp	INPUT	44

Figure 4.9: Result of Replacement Node

4.2.2 Impute Missing Value

After replacing the meaningless 0 values to missing value, the missing value need to be change into another value. This is because missing value can interrupt the learning of machine learning model. To solve this, impute node were being used to replace the missing value. The missing value will be replaced with the value of mean. Among the variable, wind speed variable has the highest percentage of missing value which is 68.71%. Since the missing value in the variable are too many, the variable wind speed will not be used.

Rejected Variables Summary
Number Of Observations

Variable Name	Label	Percent Missing
Wind_Speed__m_s_	Wind Speed (m/s)	68.7106

Figure 4.10: Rejected Variable

The value that will replace the missing value is 67.47, 1.35, 83.25, 0.036, 0.046 and 32.63 for variable PM10, CO, Humidity, NO2, O_3 and Temperature respectively:

Imputation Summary
Number Of Observations

Variable Name	Impute Method	Imputed Variable	Indicator Variable	Impute Value	Role	Measurement Level	Label
PM10_Conc	MEAN	IMP_PM10_Conc	M_PM10_Conc	67.4744	TARGET	INTERVAL	PM10_Conc
REP_Co_Conc	MEAN	IMP_REP_Co_Conc	M_REP_Co_Conc	1.3566	INPUT	INTERVAL	Replacement: Co_Conc
REP_Humidity	MEAN	IMP_REP_Humidity	M_REP_Humidity	83.2597	INPUT	INTERVAL	Replacement: Humidity
REP_No2_Conc	MEAN	IMP_REP_No2_Conc	M_REP_No2_Conc	0.0363	INPUT	INTERVAL	Replacement: No2_Conc
REP_03_Conc	MEAN	IMP_REP_03_Conc	M_REP_03_Conc	0.0469	INPUT	INTERVAL	Replacement: 03_Conc
REP_Temp	MEAN	IMP_REP_Temp	M_REP_Temp	32.6301	INPUT	INTERVAL	Replacement: Temp

Figure 4.11: Result of imputed variable

4.2.3 Transform Variable

To make classification prediction model, the target variable must be in binary class. Hence, the target variable PM10 will be classified into 2 group which are PM10 less than or equal to 50 μg/m³ and PM10 more than 50 μg/m³. The PM10 more than 50 μg/m³ will be labelled as '1' while the other group will be labelled as '0'. The transform variable conducted in excel with command:

=IF(Value>50,1,0)

The result of transformed variable can be seen in figure below:

PM10_Conc	Wind Speed (m/s)	PM10_transformed
70.000		1
58.000		1
60.000		1
43.000		0
50.000		0
47.000		0
48.000		0
48.000		0
64.000		1
58.000		1
44.000		0
47.000		0
45.000		0

Figure 4.12: Sample of transformed variable PM10

4.2.4 Drop irrelevant variable

After all the value in the variables was replaced and imputed, a drop node was used to create a new variable for each input variable. Hence, the old raw variable will be dropped to be replaced by the new variable.

Name	Label	Drop	Role	Level
Co_Conc	Co_Conc	Yes	Rejected	Interval
Date	Date	Yes	Time ID	Nominal
Humidity	Humidity	Yes	Rejected	Interval
IMP_PM10_Co	Imputed: PM10_Conc	Yes	Target	Interval
IMP_REP_Co_C	Imputed: Replacement: Co_Conc	No	Input	Interval
IMP_REP_Hum	Imputed: Replacement: Humidity	No	Input	Interval
IMP_REP_No2_	Imputed: Replacement: No2_Conc	No	Input	Interval
IMP_REP_O3_C	Imputed: Replacement: O3_Conc	No	Input	Interval
IMP_REP_Tem	Imputed: Replacement: Temp	No	Input	Interval
M_PM10_Conc	Imputation Indicator for PM10_Conc	Yes	Input	Binary
M_REP_Co_Co	Imputation Indicator for REP_Co_Conc	Yes	Input	Binary
M_REP_Humidi	Imputation Indicator for REP_Humidity	Yes	Input	Binary
M_REP_No2_C	Imputation Indicator for REP_No2_Conc	Yes	Input	Binary
M_REP_O3_Co	Imputation Indicator for REP_O3_Conc	Yes	Input	Binary
M_REP_Temp	Imputation Indicator for REP_Temp	Yes	Input	Binary
No2_Conc	No2_Conc	Yes	Rejected	Interval
O3_Conc	O3_Conc	Yes	Rejected	Interval
PM10_transfor	PM10_transformed	No	Target	Interval
Site_Id	Site_Id	Yes	ID	Nominal
Site_Location	Site_Location	Yes	Text	Nominal
So2_Conc	So2_Conc	No	Input	Interval
Temp	Temp	Yes	Rejected	Interval
Wind_Speed_	Wind Speed (m/s)	Yes	Rejected	Interval

Figure 4.13: List of Dropped Variables

4.2.5 Variable Selection

Before constructing the model, selection variable needs to be conducted to select only suitable variables to be used. Stepwise variable selection is used in this study. Through this method, only 4 input are used while another 2 have been rejected. Below is the list of accepted and rejected variables.

Name /	Use	Report	Role	Level	Model
IMP_REP_Co_Conc	Default	No	Input	Interval	
IMP_REP_Humidity	Default	No	Input	Interval	
IMP_REP_No2_Conc	Default	No	Input	Interval	
IMP_REP_O3_Conc	Default	No	Input	Interval	
IMP_REP_Temp	Default	No	Rejected	Interval	
PM10_transformed	Yes	No	Target	Nominal	Reg
So2_Conc	Default	No	Rejected	Interval	

Figure 4.14: List of Accepted and Rejected Variables

4.3 PREDICTION MODELS

4.3.1 Neural Network with Logistic Activation Function

The model constructed using AutoNeural node in SAS Enterprise Miner. The node will run models with multiple hidden neuron node and will select only the best model among them. The figure below shows the final model for Neural Network with Logistic Activation Function. The final model will have 2 hidden neurons in hidden layer.

```
Final Model
Stopping: Termination criteria were satisfied: overfitting based on _VMISC_

 _func_      _AVERR_     _VAVERR_     neurons

LOGISTIC     0.50607      0.51222        1
LOGISTIC     0.49506      0.50158        1
                                     =======
                                        2
```

Figure 4.15: Final Model for Neural Network Logistic Activation Function

The final model above can be visualized as follows:

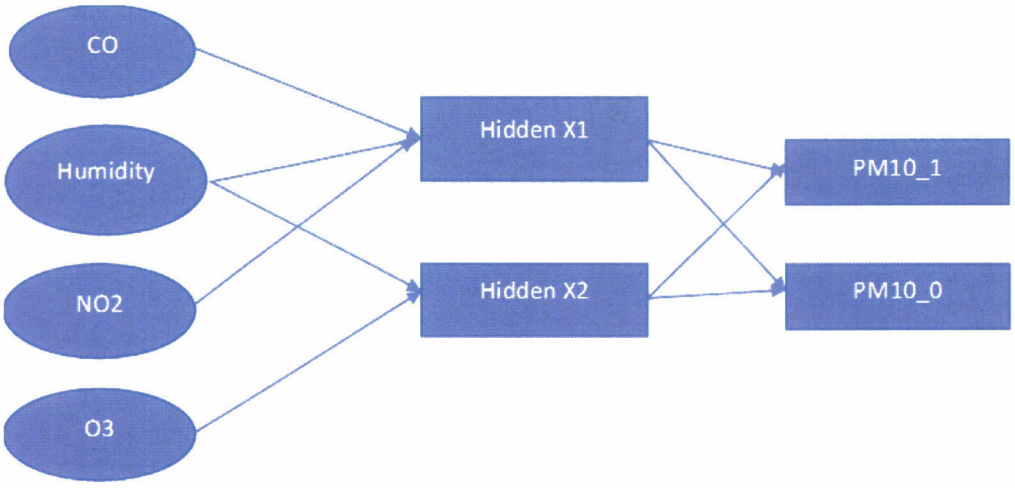


Figure 4.16: Visualization of The Neural Network Logistic Function Model

From the output of Autoneural, the performance of the models can be seen in confusion matrix, there are 2 confusion matrices, which are for training and validation dataset.

TRAINING:

Table 4.2:

Confusion Matrix of Training Set of Neural Network Logistic Activation Function

		Predicted:		
	n=1222	PM10 <= 50 μg/m³	PM10 > 50 μg/m³	
Actual:	PM10 <= 50 μg/m³	TN = 253	FP = 178	431
	PM10 > 50 μg/m³	FN = 126	TP = 665	791
		379	843	

Precision = 0.7888

Misclassification rate = 0.249

Specificity = 0.587

Sensitivity = 0.841

Accuracy = 0.536

VALIDATION:

Table 4.3:

Confusion Matrix of Validation Set of Neural Network Logistic Activation Function

		Predicted:		
	n=523	PM10 <= 50 μg/m³	PM10 > 50 μg/m³	
Actual:	PM10 <= 50 μg/m³	TN = 121	FP = 63	184
	PM10 > 50 μg/m³	FN = 55	TP =284	339
		176	347	

Precision = 0.818

Misclassification Rate = 0.225

Specificity = 0.658

Sensitivity = 0.837

Accuracy = 0.774

The value of precision is about how the model correctly predict positive result. The result shows that the value of precision increase from training to validation set. The misclassification rate decrease in validation set, which indicate the model predict more wrong class in training set. Sensitivity value indicate when the PM10 value exceed 50, how much does it predict PM10 value is indeed exceed 50. The performance indicates that the model's sensitivity is quite less compared to its performance in training set. Lastly, the overall accuracy showed that the model has the tendency to make more correct prediction in validation set compared to in training set. Based on the confusion matrices, the model's accuracy in validation set is higher compared to its accuracy in training set.

4.3.2 Neural Network with Hyperbolic Tangent Activation Function

The model also constructed by using AutoNeural node in SAS Enterprise Miner. The node will run models with multiple hidden neuron node and will select only the best model among them. The figure below shows the final model for Neural Network with Hyperbolic Tangent Activation Function. The final model will have 1 hidden neuron in the hidden layer.

```
Final Model
Stopping: Termination criteria were satisfied: overfitting based on _VMISC_

_func_     _AVERR_     _VAVERR_     neurons

 TANH      0.48789     0.50380        1
                                   =======
                                      1
```

Figure 4.17: Final Model for Neural Network Hyperbolic Tangent Activation Function

The Hyperbolic tangent Neural Network model above can be visualized as follows:

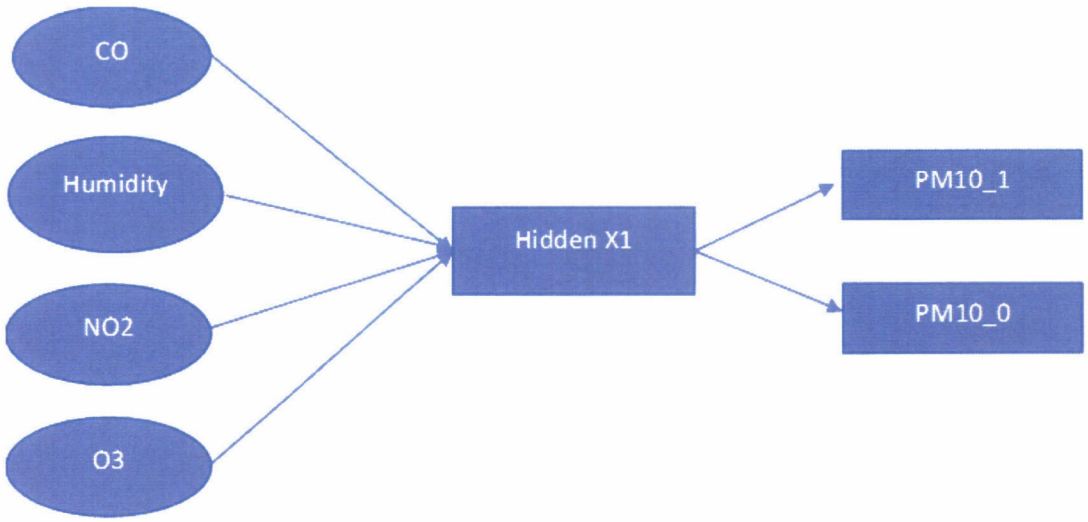


Figure 4.18: Visualization of The Neural Network Hyperbolic Tangent Function Model

From the output of Autoneural for this model, the performance of the models can be seen in confusion matrix, there are 2 confusion matrices, which are for training and validation dataset.

TRAINING:

Table 4.4:

Confusion matrix of training set of Neural Network Hyperbolic Tangent function

		Predicted:		Total
	n=1222	PM10 <= 50 μg/m³	PM10 > 50 μg/m³	
Actual:	NO	TN = 245	FP = 186	431
	YES	FN = 115	TP = 676	791
Total		540	862	

Precision = 0.784

Misclassification rate = 0.246

Specificity = 0.568

Sensitivity = 0.784

Accuracy = 0.754

VALIDATION:

Table 4.5:

Confusion matrix of validation set of Neural Network Hyperbolic Tangent function

		Predicted:		
	n=523	PM10 <= 50 μg/m³	PM10 > 50 μg/m³	
Actual:	NO	TN = 113	FP = 71	184
	YES	FN = 45	TP =294	339
		158	365	

Precision = 0.805

Misclassification Rate = 0.222

Specificity = 0.601

Sensitivity = 0.867

Accuracy = 0.778

In confusion matrix, precision indicates how the model predict positive result correctly. As shown in the calculation, the value of precision increase from training set to validation set. However, in misclassification rate, it indicates how the model predict wrong result. In this case, the model shows that there is a decrease in amount of misclassification rate from training set to validation set. Sensitivity value indicate when the PM10 value exceed 50, how much does it predict PM10 value is indeed exceed 50. The performance indicates that the model's sensitivity is more compared to its performance in training set. Lastly, for accuracy, it indicates in overall, the model has the tendency to make more correct prediction than make incorrect prediction. As shown in the confusion matrix, there is an increase in accuracy for validation dataset compared to training dataset.

4.4 MODEL COMPARISON

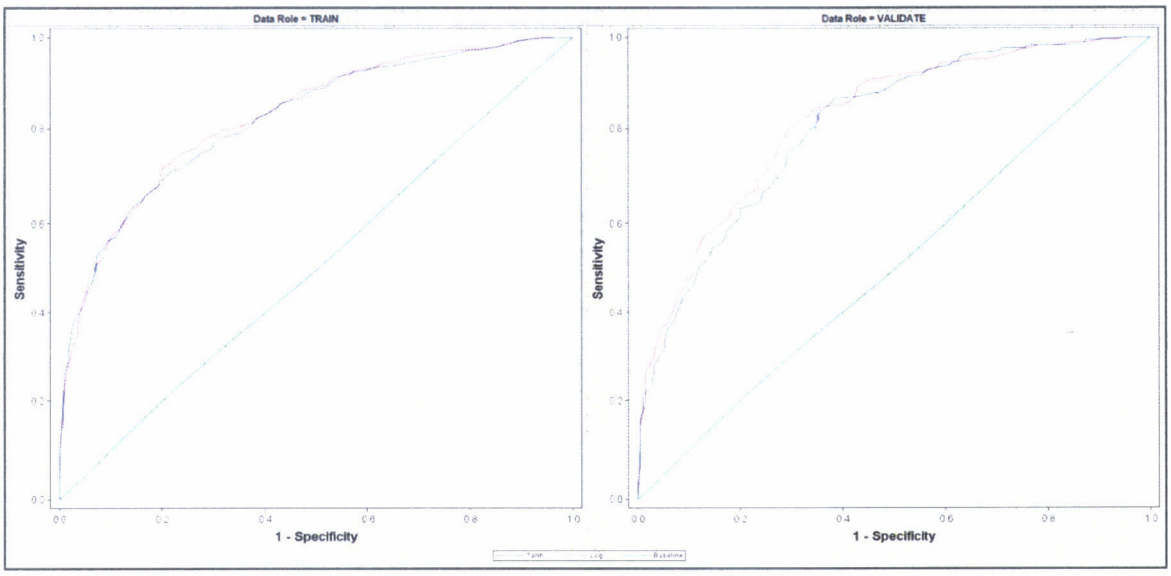


Figure 4.19: ROC Chart of Both Activation Functions

The chart above shows ROC chart comparison between those 2 models. The left chart indicates ROC chart in training set while the right one is ROC chart from validation set. The obvious pattern that can be seen is the performance of both models is on par in training dataset. However, in validation set, Logistic Neural Network shows slightly increment in term of the area below the ROC curve. The ROC curve is an indicator of the model's performance in predicting the target variable. Since Logistic Neural Network have slightly bigger ROC curve than Hyperbolic tangent Neural Network in validation set, Logistic Neural Network have higher performance than Hyperbolic Tangent Neural Network.

Table 4.6:

Model Comparison of Both Activation Functions

Model Description	Tanh	Log
Train: Misclassification Rate	0.24631751	0.2487725
Valid: Misclassification Rate	0.22179732	0.22562142
Gap	-0.025	-0.023
Train: Average Squared Error	0.16281294	0.16090932
Valid: Average Squared Error	0.16688717	0.16220321
Gap	0.004074	0.001294
Train: Roc Index	0.822	0.827
Valid: Roc Index	0.805	0.819
Gap	-0.017	-0.008

Based on the table above, Neural Network with Logistic activation function perform much better than Neural Network with Hyperbolic tangent function. The table shows that Logistic Neural Network has the lowest gap of misclassification rate which mean this model have the tendency to make correct classification rather than making wrong classification. The model's average square error also smaller compared to the Hyperbolic Tangent function. Meanwhile Hyperbolic tangent Neural Network have the highest ROC index. Using those 3 criteria, Neural Network with Logistic activation function is considered as the best model since it outshines 2 of 3 criteria compared to the other.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATIONS

The main objective of this study is to determine the best activation function to be used in the model to predict the PM10 concentration in Klang, Selangor. Various activation function in MLP networks had been investigated to determine the most suitable model to solve that particular problem. The outcome obtained from each network were compared to determine the most suitable activation function to be used. Not only that, this study also will identify variables that have high correlation in predicting the output which is PM10 concentration.

Among the variable, Carbon Monoxide give the highest correlation of 0.576 compared to other variable in predicting PM10 concentration. Hence, in selecting variable to be used in predicting model, Carbon Monoxide has been selected to be one of the variables. Aside from Carbon Monoxide, the other variables which are humidity, Nitrogen Dioxide and Ozone also has been used. The selected variable has been done using stepwise selection method.

In determining the best model, 2 activation function has been used which are Logistic function and Hyperbolic Tangent function. The 3 criteria to determine the best model are the lowest difference between training set and validation set in misclassification rate and average square error and the highest difference between training set and validation set in ROC index. Logistic function has the lowest difference between training and validation set misclassification rate and Hyperbolic tangent function has the highest difference in ROC index. Since Logistic function has met two out of these three criteria, this paper conclude that Logistic Activation Function is the best model to predict the PM10 concentration.

Hence, it is recommended that for future research regarding prediction in air quality to use Multi-layer perceptron with Logistic Activation function since it is proven that multi-layer perceptron with this activation function give less error and misclassification rate while giving higher performance compared to Multi-Layer perceptron with Hyperbolic tangent function.

REFERENCES

Abdullah, S, Ismail, M., & Ahmed, A. N. (2020). Multi-Layer Perceptron Model for Air Quality Prediction. *Malaysian Journal of Mathematical Sciences*, *13(S)* (0 SE-Articles). http://einspem.upm.edu.my/jurnal/index.php/mjms/article/view/530

Abdullah, Samsuri, Ismail, M., & Fong, S. Y. (2017). Multiple Linear Regression (MLR) models for long term Pm 10 concentration forecasting during different monsoon seasons. *Journal of Sustainability Science and Management*, *12*(1), 60–69.

Ahmad, R. H. (2019, August 29). *We're unaware that air pollution is a silent killer*. NST Online. https://www.nst.com.my/opinion/letters/2019/08/516887/were-unaware-air-pollution-silent-killer#:%7E:text=According%20to%20a%20Universiti%20Malaya,leading%20to%20death%20and%20diseases.

Barmpadimos, I., Hueglin, C., Keller, J., Henne, S., & Prévôt, A. S. H. (2011). Influence of meteorology on PM10 trends and variability in Switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics*, *11*(4), 1813–1835. https://doi.org/10.5194/acp-11-1813-2011

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G., & Di Carlo, P. (2017). Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollution Research*, *8*(4), 652–659. https://doi.org/10.1016/j.apr.2016.12.014

Braniš, M., & Větvička, J. (2010). PM10, Ambient temperature and relative humidity during the XXIX summer olympic games in Beijing: Were the athletes at risk? *Aerosol and Air Quality Research*, *10*(2), 102–110. https://doi.org/10.4209/aaqr.2009.09.0055

Buteau, S., & Goldberg, M. S. (2016). A structured review of panel studies used to investigate associations between ambient air pollution and heart rate variability. *Environmental Research*, *148*(July 2016), 207–247. https://doi.org/10.1016/j.envres.2016.03.013

Dedovic, M. M., Avdakovic, S., Turkovic, I., Dautbasic, N., & Konjic, T. (2016). Forecasting PM10 concentrations using neural networks and system for improving air quality. *2016 11th International Symposium on Telecommunications, BIHTEL 2016, March 2019*. https://doi.org/10.1109/BIHTEL.2016.7775721

Flach, P., Hernández-Orallo, J., & Ferri, C. (2011). A coherent interpretation of AUC as a measure of aggregated classification performance. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, 657–664.

García Nieto, P. J., Martínez Torres, J., Araújo Fernández, M., & Ordóñez Galán, C. (2012). Support vector machines and neural networks used to evaluate paper manufactured using Eucalyptus globulus. *Applied Mathematical Modelling*, *36*(12), 6137–6145. https://doi.org/10.1016/j.apm.2012.02.016

García Nieto, P. J., Sánchez Lasheras, F., García-Gonzalo, E., & de Cos Juez, F. J. (2018). PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study. *Science of the Total Environment*, *621*, 753–761. https://doi.org/10.1016/j.scitotenv.2017.11.291

Harrison, R. M., Deacon, A. R., Jones, M. R., & Appleby, R. S. (1997). Sources and processes affection concentrations of PM10 and PM2.5 particulate matter in Birmingham (U.K.). *Atmospheric Environment*, *31*(24), 4103–4117. https://doi.org/10.1016/S1352-2310(97)00296-3

Isa, I. S., Saad, Z., Omar, S., Osman, M. K., Ahmad, K. A., & Sakim, H. A. M. (2010). Suitable MLP network activation functions for breast cancer and thyroid disease detection. *Proceedings - 2nd International Conference on Computational Intelligence, Modelling and Simulation, CIMSim 2010, June 2014*, 39–44. https://doi.org/10.1109/CIMSiM.2010.93

Latif, M. T., Abidin, E. Z., & Praveena, S. M. (2015). The Assessment of Ambient Air Pollution Trend in Klang Valley. *World Environment*, *5*(1), 1–11. https://doi.org/10.5923/j.env.20150501.01

Lešnik, U., Mongus, D., & Jesenko, D. (2019). Predictive analytics of PM10 concentration levels using detailed traffic data. *Transportation Research Part D: Transport and Environment*, *67*(December 2018), 131–141. https://doi.org/10.1016/j.trd.2018.11.015

Ling, T., Lewellen, T. K., & Miyaoka, R. S. (2007). Depth of interaction decoding of a continuous crystal detector module. *Physics in Medicine and Biology*, *52*(8), 2213–2228. https://doi.org/10.1088/0031-9155/52/8/012

McKendry, I. G. (2002). Evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. *Journal of the Air and Waste Management Association*, *52*(9), 1096–1101. https://doi.org/10.1080/10473289.2002.10470836

Park, S., Kim, M., Kim, M., Namgung, H. G., Kim, K. T., Cho, K. H., & Kwon, S. B. (2018). Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN). *Journal of Hazardous Materials*, *341*, 75–82. https://doi.org/10.1016/j.jhazmat.2017.07.050

Patra, A. K., Gautam, S., Majumdar, S., & Kumar, P. (2016). Prediction of particulate matter concentration profile in an opencast copper mine in India using an artificial neural network model. *Air Quality, Atmosphere and Health*, *9*(6), 697–711. https://doi.org/10.1007/s11869-015-0369-9

Pawul, M., & Śliwka, M. (2016). Application of artificial neural networks for prediction of air pollution levels in environmental monitoring. *Journal of Ecological Engineering*, *17*(4), 190–196. https://doi.org/10.12911/22998993/64828

Querol, X., Alastuey, A., Ruiz, C. R., Artiñano, B., Hansson, H. C., Harrison, R. M., Buringh, E., Ten Brink, H. M., Lutz, M., Bruckmann, P., Straehl, P., & Schneider,

J. (2004). Speciation and origin of PM10 and PM2.5 in selected European cities. *Atmospheric Environment*, *38*(38), 6547–6555. https://doi.org/10.1016/j.atmosenv.2004.08.037

Russo, A., Lind, P. G., Raischel, F., Trigo, R., & Mendes, M. (2015). Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales. *Atmospheric Pollution Research*, *6*(3), 540–549. https://doi.org/10.5094/APR.2015.060

Sayegh, A. S., Munir, S., & Habeebullah, T. M. (2014). Comparing the performance of statistical models for predicting PM10 concentrations. *Aerosol and Air Quality Research*, *14*(3), 653–665. https://doi.org/10.4209/aaqr.2013.07.0259

Schornobay-Lui, E., Alexandrina, E. C., Aguiar, M. L., Hanisch, W. S., Corrêa, E. M., & Corrêa, N. A. (2019). Prediction of short and medium term PM 10 concentration using artificial neural networks. *Management of Environmental Quality: An International Journal*, *30*(2), 414–436. https://doi.org/10.1108/MEQ-03-2018-0055

Shakerkhatibi, M., Mohammadi, N., Zoroufchi Benis, K., Sarand, A. B., Fatehifar, E., & Hashemi, A. A. (2015). Using ANN and EPR models to predict carbon monoxide concentrations in urban area of Tabriz. *Environmental Health Engineering and Management Journal*, *2*(3), 117–122.

Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., de` Donato, F., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F., & Kloog, I. (2017). Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environment International*, *99*(2017), 234–244. https://doi.org/10.1016/j.envint.2016.11.024

Study, C., Gocheva-ilieva, S. G., & Stoimenova, M. P. (2018). *PM10 Prediction and Forecasting Using CART : A. 12*(9), 572–577.

Tadeusiewicz, R., & Dobrowolski, J. W. (2004). Artificial intelligence and primary prevention of health hazard related to changes of elements in the environment. *Polish Journal of Environmental Studies*, *13*(3), 349–352.

Ul-Saufie, A. Z., Yahaya, A. S., Ramli, N. A., & Hamid, H. A. (2015). PM10 concentrations short term prediction using feedforward backpropagation and general regression neural network in a sub-urban area. In *Journal of Environmental Science and Technology* (Vol. 8, Issue 2, pp. 59–73). https://doi.org/10.3923/jest.2015.59.73

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

World Health Organization. (2006). *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment* (No. WHO/SDE/PHE/OEH/06.02). World Health Organization.

Yin, J., & Tian, L. (2014). Joint confidence region estimation for area under ROC curve and Youden index. Statistics in Medicine, 33(6), 985–1000. https://doi.org/10.1002/sim.5992

Zhou, X., Wang, X., Hu, C., & Wang, R. (2020). An analysis on the relationship between uncertainty and misclassification rate of classifiers. *Information Sciences*, *535*, 16–27. https://doi.org/10.1016/j.ins.2020.05.059

Zolkepli, F. (2020, March 10). Air pollution: An underestimated killer. The Star Online. https://www.thestar.com.my/lifestyle/health/2020/03/10/air-pollution-an-underestimated-killer

APPENDICES

APPENDIX A

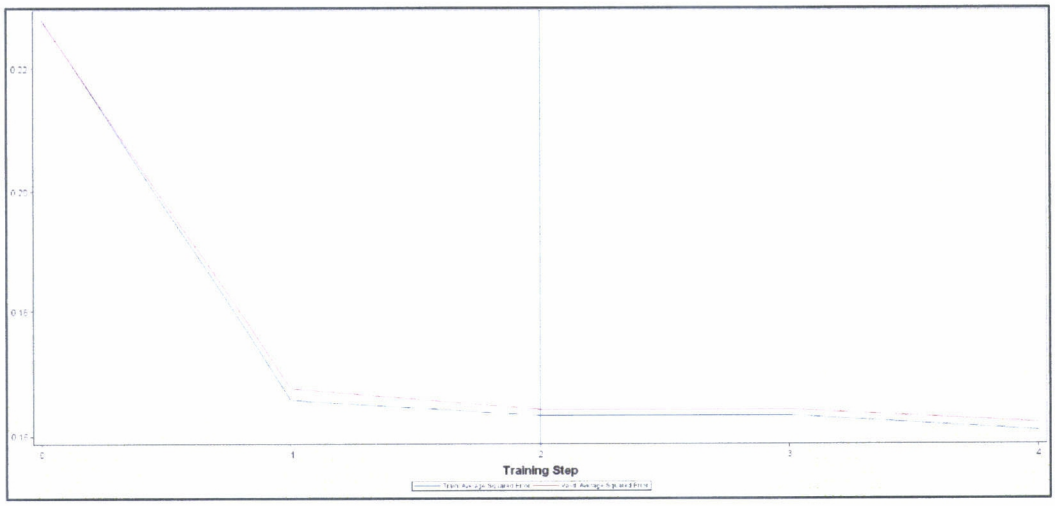
Description of The Model with Logistic Activation Function

1. Final Training History of the model with Logistic Activation Function.

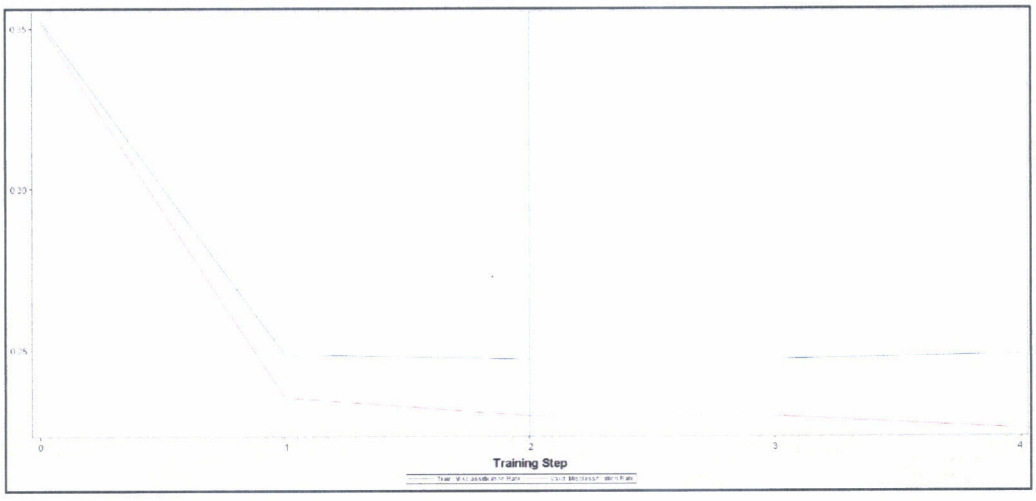
Final Training History

step	_func_	_status_	_iter_	_AVERR_	_MISC_	_AIC_	_VAVERR_	_VMISC_
SINGLE LAYER 1	LOGISTIC	initial	0	0.64910	0.35270	1600.41	0.64857	0.35182
SINGLE LAYER 1	LOGISTIC	keep	2	0.50607	0.24877	1250.84	0.51222	0.23518
SINGLE LAYER 2	LOGISTIC	keep	2	0.49506	0.24714	1235.93	0.50158	0.22945
SINGLE LAYER 3	LOGISTIC	reject	0	0.49506	0.24714	1247.93	0.50158	0.22945
		Final	5	0.48450	0.24877	1210.12	0.49338	0.22562

2. Average Square Error graph of the model with Logistic Activation Function.



3. Misclassification Rate graph of the model with Logistic Activation Function.



4. Confusion Matrix of the model with Logistic Activation Function.

Event Classification Table

Data Role=TRAIN Target=PM10_transformed Target Label=PM10_transformed

False Negative	True Negative	False Positive	True Positive
126	253	178	665

Data Role=VALIDATE Target=PM10_transformed Target Label=PM10_transformed

False Negative	True Negative	False Positive	True Positive
55	121	63	284

APPENDIX B

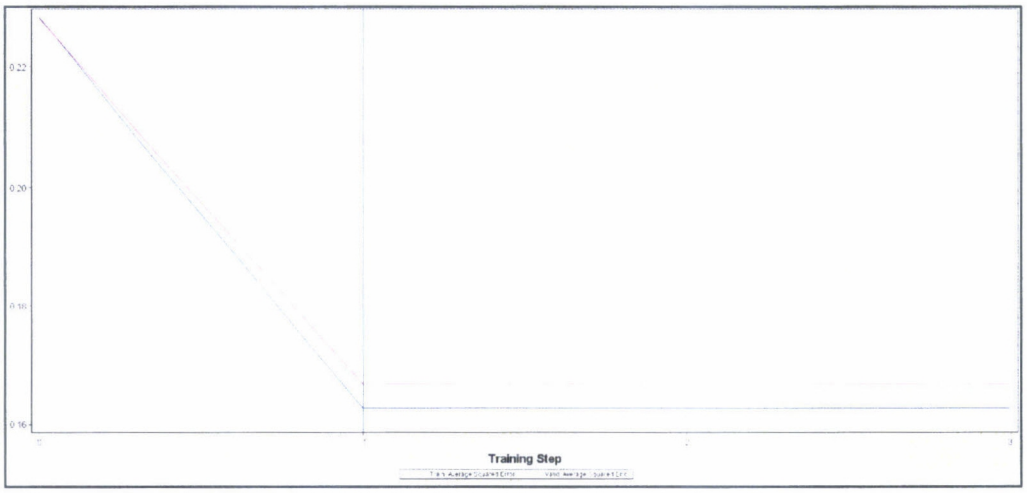
Description of The Model with Hyperbolic Tangent Function

1. Final Training History of the model with Hyperbolic Tangent Activation Function.

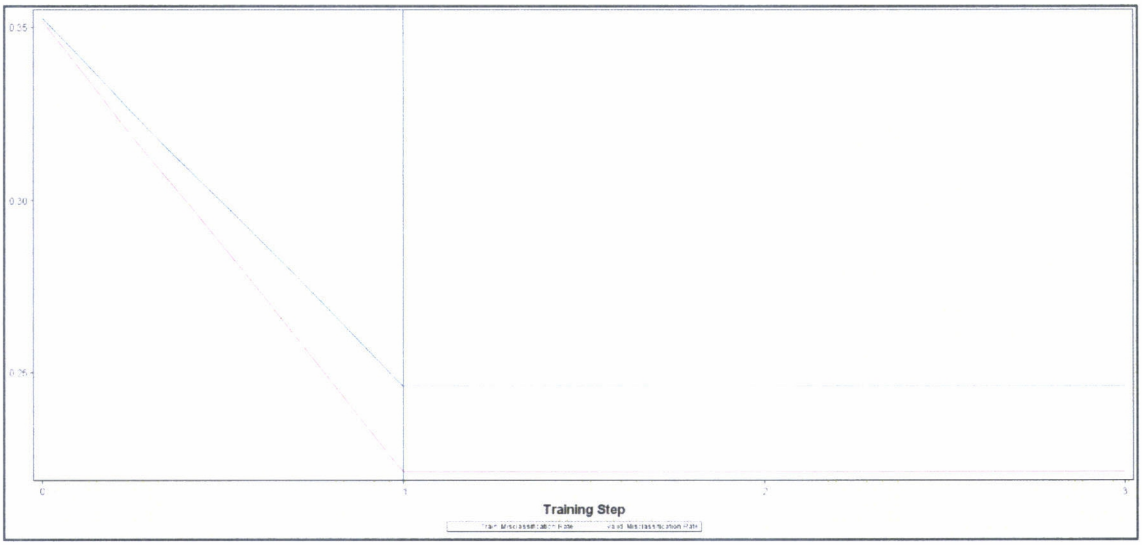
Final Training History

step	_func_	_status_	_iter_	_AVERR_	_MISC_	_AIC_	_VAVERR_	_VMISC_
SINGLE LAYER 1	TANH	initial	0	0.64910	0.35270	1600.41	0.64857	0.35182
SINGLE LAYER 1	TANH	keep	3	0.48789	0.24632	1206.41	0.50380	0.22180
SINGLE LAYER 2	TANH	reject	0	0.48789	0.24632	1218.41	0.50380	0.22180
		Final	0	0.48789	0.24632	1206.41	0.50380	0.22180

2. Average Square Error graph of the model with Hyperbolic Tangent Activation Function.



3. Misclassification Rate graph of the model with Hyperbolic Tangent Activation Function.



4. Confusion Matrix of the model with Hyperbolic Tangent Activation Function.

Event Classification Table

Data Role=TRAIN Target=PM10_transformed Target Label=PM10_transformed

False Negative	True Negative	False Positive	True Positive
115	245	186	676

Data Role=VALIDATE Target=PM10_transformed Target Label=PM10_transformed

False Negative	True Negative	False Positive	True Positive
45	113	71	294