Implementation of Neural Network Method for Air Quality Forecasting in Jakarta Region

Dinar Ajeng Kristiyanti1, Esty Purwaningsih2, Ela Nurelasari2, Ahmad Al Kaafari2, and Akhmad Hairul Umam3

1Informatics Engineering, STMIK Nusa Mandiri, Indonesia
2Information System, Bina Sarana Informatika University, Indonesia
3Communication, Tanri Abeng University, Indonesia
E-mail: dinar@nusamandiri.ac.id

Abstract. The quality of air can be influenced by the amount of pollution that occurs in an area. The city of Jakarta is ranked in the top ten as the nation’s capital with the worst air quality in the world. Poor air quality both inside and outside the room can have an impact on the emergence of various diseases and even death. For this reason, forecasting of air quality in the city of Jakarta, Indonesia is needed to anticipate the likely impact that will arise. In this study forecasting air quality using the neural network method in which this method has the advantage of being able to solve problems, especially large data samples and has been able to prove in handling non-linear problems. The data collection used is secondary data from the Environmental Service Office of DKI Jakarta Province as many as 2989 records with variables as determinants consisting of 5 of which PM10, SO2, CO, O3, NO2 and 1 output variable are good, moderate, unhealthy and very unhealthy. From the calculations result in this study it is known that the Neural Network method obtained an accuracy performance of 88.86% in which the Lubang Buaya area noted as the most unhealthy air quality.

1. Introduction
Air pollution is perceived as an important environmental problem that almost happens in all countries and needs to be considered seriously. Changes in the characteristics of the atmosphere are caused by the amount of environmental contamination both inside and outside the room that contains chemicals. Smoke fumes from households, vehicle exhausts, fires and industry are causes of air pollution [1][2]. The emergence of critical and chronic diseases for humans is caused by poor air [3]. According to the latest data from WHO, every year 4,200,00 people die from air pollution and 91% of the world population still lives in areas that air quality is above the WHO standards [4].

The impact of air quality both inside and outside the room can cause death more quickly with a range of 7 (seven) million per year caused by various diseases such as stroke, heart, chronic obstructive pulmonary, lung cancer, respiratory infections, etc. More than 80% of people living in urban areas with very high level of air pollution and low and middle income countries are the most highlighted categories [2]. Particles (PM), carbon monoxide (CO), ozone (O3), nitrogen dioxide (NO2) and sulphur dioxide (SO2) are considered as the most dangerous pollutants for public health [5]. The most promising alternative method is Artificial neural network (ANN)
This model is often used for forecasting air quality, especially for forecasting every hour [8].

In 2017 to 2018, in the Jakarta city region the level of PM2.5 concentration at an average level increased by more than 50%, from 20 µg/m³. By 2019 the pollution level had exceeded 2018. Poor air quality in the Jakarta region was largely the result of the process of rapid air growth and development, coupled with the availability of agriculture, coal-fired power plants and immediate vehicles. With the increase in the city it is feared that it can affect the air quality resulting from construction projects that produce dust, as well as the ever-increasing contribution that traffic jams can occur. [9]. In 2018, the city of Jakarta was ranked in the top ten as the capital of the country with the worst air quality in the world, where the city of Jakarta reached four times above the annual safe limit according to World Health Organization (WHO) standards, which is 10 µg / m³. This figure has also far exceeded the annual safe limit according to national standards in PP No. 41 of 1999 concerning Air Pollution Control, which is 15 µg/m³[10]. So, there is a need to do a research study to determine the prediction model toward the level of air pollution in Jakarta so that forecasting can be done for the future in every day. Air quality data in the Jakarta region is a time series stochastic process so that it is possible to make historical data forecasting to predict the future every day in the hope that the right policy can be taken as a step to anticipate.

The ANN is a modelling technique that can determine the non-linear relationship between input variables and output variables. It is based on the training process that can estimate the value of the output variable for the input dataset. It also requires an adequate amount of data for training; after the training is satisfactory, it must be able to provide outputs for inputs that were previously invisible [11][12]. Neural networks have the ability to complex models with a variety of nonlinear problems, but the main disadvantage of neural networks is their instability, especially in conditions of noise and limited datasets [13].

Several studies have shown that the neural network model can achieve good accuracy [14]. There are many studies have been carried out for prediction of air quality [15]. To measure the increase in pollution, air quality forecasting techniques have been rapidly upgraded according to demand. Models such as recurrent neural networks (RNN) and short-term memory (LSTM) are applied to perform meteorological forecasting, weather forecasting [16], estimation of air pollution and probability of rainfall [17]. The Neural Network model has been used recently in many studies to estimate air pollutant emissions [18], which in this study illustrates the development of the Artificial Neural Network model for calculating road emissions for PM10 and four other pollutants. Research [19], Zuoying District in Taiwan has the worst air pollution due to the high PM 2.5 concentration in the region and hence studies of air quality in this region are very important. New approach proposed by BRNN/FFS in developing PM 2.5. Studies conducted to evaluate the performance of the BRNN/FFS system show that the proposed method has reached the lowest MSE, MAE, and RMSE than other estimation methods and the resulting R-squared value shows good. Air quality warning system modeling is carried out based on the SVM approach to achieve high efficiency and fast response, which is an important factor when we consider warning systems. In this study, the accurate results have been obtained for the BRNN/FFS estimation and SVM classifier which show that the proposed air quality warning system has proven to be efficient.

Method of Neural network has been employed in various areas of research. In particular, it has been used to predict the wind speed and atmospheric ozone concentration [20][21][22]. However, method of Neural network is rarely used as an option to link the Meteorological modelling result. This research is aimed at developing a neural network based approach which is known as the Ensemble model to increase the accuracy of the meteorological input fields used in modelling of air quality. In a case study [23] that research on air pollution that occurred in North China in the January-October 2006 period was carried out through the model of MM5,
WRF, and Ensemble. It was found that Ensemble model had higher simulation accuracy than MM5 and WRF models. When we did a comparison with MM5eCMAQ and WRFeCMAQ model, the performance result of CMAQ ensemble model was increased from 7.0% to 11.1% and from 17.8% to 27.5%. Generally, neural network-based Ensemble model is effective options for linking Meteorological modelling result. It is also associated with less computational costs compared to model of numerical Meteorological. So, it can be predicted that the proposed neural network based approach can also be used to other areas to increase the effectiveness of management for air quality.

The research study was conducted by applying neural network method for forecasting air quality in the Jakarta region, where the neural network method is known to be able to solve problems, especially large data samples obtained from the Jakarta Provincial Environment Office and to improve the accuracy of an algorithm and greater data classification. This study has the aim of knowing the value of accuracy and can produce output which is the predicted value by using the Neural Network method. In this study a method test was conducted using Rapid Miner tool with Neural Network method for forecasting air quality in the Jakarta region. Rapid miners have the ease of use. It is used as a tool in measuring the accuracy of experimental data conducted in research [24]. Its usefulness is to process data and assist in assessing the accuracy of the training data conducted in research [25]. In addition, to find out the level of accuracy, Confusion Matrix and RMSE (Root Mean Square Error) are used to determine the error level.

2. Method

The Neural Network method was chosen in this study because it has a very easily accepted ability for noisy data, has a higher level of accuracy and better data mining. It is also considered as one of the methods that can predict the air quality in the Jakarta region per day in the future. Research is an activity that aims to make original contributions to science. There are four types commonly used as research methods, including: action research, experimentation, case study and survey [26]. This study uses experimental research, which is research that involves investigation of parameters or variables depending on the researcher and uses tests that are controlled by the researcher himself.

The classification algorithm will produce a set of rules called rules which will be used as indicators to be able to predict the class of data you want to predict [26]. In this air quality forecasting research, we conducted the method testing phase with the following steps:

a. The data collection used is secondary data obtained from the Jakarta Environmental Department as many as 2989 records, with variables as determinants consisting of 5 of which are pm10, so2, co, o3, no2 and 1 output variable that is good, medium, not healthy and very unhealthy. Preprocessing data is done by replacing the missing value record with an average value

b. The dataset is transformed into a range of 0 to 1. Then the dataset is divided by the 10-fold cross validation method, which is divided into testing data and training data. Then the data is processed and tested by Rapid Miner using the Neural Network (NN) method. The neural network process will continue to repeat as many as 10 repetitions.

c. Furthermore, this research produces an accuracy level, namely the Confusion Matrix to compare the accuracy value and the RMSE value to determine the error rate in the Neural Network method.

3. Result and Discussion

This research uses neural network method by dividing the dataset through 10-fold cross validation, which is divided into testing data and training data. The dataset in this study was obtained from air condition data according to air quality monitoring stations (SPKU) in DKI Jakarta Province which had attributes including: particulate matter (pm10), sulphur (so2),
Table 1. National ambient air quality standard according to PP No. 41 of 1999

| No | Parameter                | Time  | Quality Standard |
|----|--------------------------|-------|------------------|
| 1  | Aerosol/Particulate (PM10) | 24 Hours | 150 µg/m³  |
| 2  | Carbon monoxide (CO)    | 1 Hour | 30000 µg/m³    |
|    |                          | 24 Hours | 10000 µg/m³   |
| 3  | Ozone (O3)              | 1 Hour | 235 µg/m³      |
|    |                          | 1 Year  | 50 µg/m³      |
| 4  | Sulphur dioxide (SO2)   | 24 Hour | 365 µg/m³     |
|    |                          | 1 Year  | 80 µg/m³      |
| 5  | Nitrogen dioxide (NO2)  | 1 Hour  | 0.25 µg/m³    |
|    |                          | 1 Year  | 100 µg/m³     |

Table 2. Neural Network Experiment Results (air quality dataset)

| Parameter | Neural Network | Accuracy | RMSE |
|-----------|----------------|----------|------|
| Training Cycles | Learning Rate | Momentum | Hidden Layer 1 |       |       |
| 500       | 0.3            | 0.2      | 2        | 77.32% | 0.422 |
| 500       | 0.3            | 0.2      | 6        | 88.06% | 0.323 |
| 500       | 0.1            | 0.1      | 6        | 88.52% | 0.327 |
| 500       | 0.1            | 0.2      | 6        | 88.52% | 0.327 |
| 300       | 0.1            | 0.2      | 6        | 88.42% | 0.329 |
| 450       | 0.3            | 0.2      | 6        | 87.82% | 0.324 |
| 1000      | 0.1            | 0.2      | 12       | 88.86% | 0.320 |
| 1000      | 0.1            | 0.2      | 14       | 88.56% | 0.325 |

carbon monoxide (CO), ozone (O3), nitrite (NO2) and output: good, medium, unhealthy and very unhealthy. National ambient air quality standards are set as a maximum limit for ambient air quality to prevent air pollution [28]. To protect public health and comfort, the government sets national ambient air quality standards. This can be seen in the table 1.

After the experiments conducted with neural networks, the result of accuracy and RMSE can be seen on the neural network method from 8 experiments on the air quality dataset. The neural net that results from the processing of air quality testing data by the neural network method is the multilayer perceptron. The resulting neural net uses three layers consisting of an input layer consisting of five vertices including PM10, SO2, CO, O3, NO2 and one bias node. The second layer is a hidden layer which consists of 12 nodes and one bias node. The third layer is the output layer there are four vertices which represent good, medium, unhealthy and very unhealthy classes.

The learning rate, momentum, training cycle and input layer parameters are taken from each of the best values. Several experiments were conducted. The results used as initial reference are those whose error values are lower than others and accuracy values that are higher than others. In this study, using parameter rules consisting of training cycles, learning rate as the speed of learning the system to determine the weight of neurons, momentum as a constant momentum and the number of hidden as the number of hidden layers.

In Table 2 it can be seen that from 8 experiments showing that the learning rate of 0.1, momentum 0.2 and neuron 12 produces a higher accuracy rate of 88.86% and has a lower error rate than the others which is worth 0.320.
Data testing is done through a method of Neural Network which shows that predictions created by using the method of Neural Network. In addition, from the data itself can be determined that the predictions and actual results are the same value, meaning that the Neural Network method can predict classes one day ahead for all instances precisely.

### Table 3. Air Quality Prediction Results.

| Station                      | PM10 | SO2  | CO  | O3  | NO2 | Class | Prediction |
|------------------------------|------|------|-----|-----|-----|-------|------------|
| DKI I (Bunderan HI)          | 24.0 | 17.0 | 16.0| 29.0| 2.0 | Good  | Good       |
| DKI I (Bunderan HI)          | 41.0 | 17.0 | 22.0| 23.0| 5.0 | Good  | Good       |
| DKI 2 (Kelapa Gading)        | 37.0 | 3.0  | 19.0| 52.0| 12.0| Medium | Medium     |
| DKI 2 (Kelapa Gading)        | 37.0 | 38.0 | 38.0| 89.0| 4.0 | Medium | Medium     |
| DKI 3 (Jagakarsa)            | 8.0  | 22.0 | 21.0| 7.0 | 4.0 | Good  | Good       |
| DKI 3 (Jagakarsa)            | 11.0 | 22.0 | 2.0 | 1.0 | 3.0 | Good  | Good       |
| DKI 4 (Lubang Buaya)         | 76.0 | 17.0 | 31.0| 155.0|12.0| Unhealthy | Unhealthy |
| DKI 4 (Lubang Buaya)         | 82.0 | 3.0  | 14.0| 72.0| 13.0| Medium | Medium     |
| DKI 5 (Kebon Jeruk) West Jakarta | 54.0| 17.0 | 21.0| 93.0| 9.0 | Medium | Medium     |
| DKI 5 (Kebon Jeruk) West Jakarta | 73.0| 3.0  | 14.0| 65.0| 8.0 | Medium | Medium     |

### 4. Conclusion and Suggestion

After testing the data on air quality forecasting in the Jakarta region using the Neural Network, it is found that the setting of learning rate 0.1, momentum 0.2 and neuron 12, results in a higher accuracy rate of 88.86% and has an error rate of 0.320 which is lower than other testing parameters and from the predicted and actual results produce the same value, which means that method of Neural Network can predict the class for the next day correctly. Based on 5 (five) stations tested in the city of Jakarta, the Lubang Buaya area is the area with the most unhealthy air quality. The Kelapa Gading and Kebon Jeruk areas (West Jakarta) have moderate air quality. The Bundaran HI and Jagakarsa areas have healthy air quality. And this is consistent with real data and predictive testing results. Therefore, it can be summarised that the Neural Network in forecasting air quality shows good and accurate result.

However, study of air quality forecasting is only limited to the Jakarta region, Indonesia. It is expected that further research can cover several cities, countries and even continents in order to obtain varied data and can use other methods besides neural networks or improvement of neural network method with other methods to improve accuracy and speed in forecasting. In further research, air quality comparison in the era of Covid-19, pre and post of Covid-19 can be compared. So that the effect or impact of Covid-19 on air quality can be obtained.

### References

[1] Jensen S S et al., 2017 High resolution multi-scale air quality modelling for all streets in Denmark Transp. Res. Part D Transp. Environ. 52 p. 322–339.
[2] Walter J, 2018, Air pollution and health: Summary. WHO.
[3] Ning X Ji X Li G and Sang N, 2019 Ambient PM2.5 causes lung injuries and coupled energy metabolic disorder Ecotoxicol. Environ. Saf. 170, December 2018 p. 620–626.
[4] Jia J Yuan X Peng X and Yan B, 2019 Cr(VI)/Pb 2+ are responsible for PM2.5-induced cytotoxicity in A549 cells while pulmonary surfactant alleviates such toxicity Ecotoxicol. Environ. Saf. 172, December 2018 p. 152–158.
[5] Ajdour A Leghibri R Chaoufi J Chirmata A Menu L and Mailler S, 2020 Towards air quality modeling in Agadir City (Morocco) Mater. Today Proc. 24, xxxx p. 17–23.
[6] Lal B and Tripathy S S, 2012 Prediction of dust concentration in open cast coal mine using artificial neural network Atmos. Pollut. Res. 3, 2 p. 211–218.

[7] Nejadkoorki F and Baroutian S, 2014 Forecasting Extreme PM10 Concentrations Using Artificial Neural Networks Int. J. Environ. Resour. 8, 1 p. 157–164.

[8] Kolehmainen M Martikainen H and Ruuskanen J, 2001 Neural networks and periodic components used in air quality forecasting Atmos. Environ. 35, 5 p. 815–825.

[9] 2020, AIR QUALITY ANALYSIS AND STATISTICS FOR JAKARTA, https://www.iqair.com/.

[10] 2020, Air quality and pollution city ranking, July 2020, https://www.iqair.com/.

[11] Palani S Lioung S Y and Tsalich P, 2008 An ANN application for water quality forecasting Mar. Pollut. Bull. 56, 9 p. 1586–1597.

[12] Voukantisis D Karatzas K Kukkonen J Räsänen T Karppinen A and Kolehmainen M, 2011 Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki Sci. Total Environ. 409, 7 p. 1266–1276

[13] Dimopoulos I F Tsiros I X Serelis K and Chronopoulou A, 2004 Combining neural network models to predict spatial patterns of airborne pollutant accumulation in soils around an industrial point emission source J. Air Waste Manag. Assoc. 54, 12 p. 1506–1515.

[14] Sözen A Güleven Z and Arcaklioğlu E, 2009 Estimation of GHG emissions in turkey using energy and economic indicators Energy Sources, Part A Recover. Util. Environ. Eff. 31, 13 p. 1141–1159.

[15] Athira V Geetha P Vinayakumar R and Soman K P, 2018 DeepAirNet: Applying Recurrent Networks for Air Quality Prediction Procedia Comput. Sci. 132 p. 1394–1403.

[16] Ghaderi A Sanandaji B M and Ghaderi F, 2017 Deep Forecast: Deep Learning-based Spatio-Temporal Forecasting ii.

[17] Lin S W Shiu Y R Chen S C and Cheng H M, 2009 Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks Expert Syst. Appl. 36, 9 p. 11543–11551.

[18] Kassomenos P Karakitsios S and Papaloukas C, 2006 Estimation of daily traffic emissions in a South-European urban agglomeration during a weekday. Evaluation of several “what if” scenarios Sci. Total Environ. 370, 2–3 p. 480–490.

[19] Balram D Lian K Y and Sebastian N, 2019 Air quality warning system based on a localized PM2.5 soft sensor using a novel approach of Bayesian regularized neural network via forward feature selection Ecotoxicol. Environ. Saf. 182, June p. 109386.

[20] Abdul-Wahab S A and Al-Alawi S M, 2002 Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks Environ. Model. Softw. 17, 3 p. 219–228.

[21] Li G and Shi J, 2010 On comparing three artificial neural networks for wind speed forecasting Appl. Energy 87, 7 p. 2313–2320.

[22] Li G Shi J and Zhou J, 2011 Bayesian adaptive combination of short-term wind speed forecasts from neural network models Renew. Energy 36, 1 p. 352–359.

[23] Cheng S Li L Chen D and Li J, 2012 A neural network based ensemble approach for improving the accuracy of meteorological fields used for regional air quality modeling J. Environ. Manage. 112 p. 404–414.

[24] Kristiyanti D A and Wahyudi M, 2017 Feature selection based on Genetic algorithm, particle swarm optimization and principal component analysis for opinion mining cosmetic product review in 2017 5th International Conference on Cyber and IT Service Management, CITSM 2017.

[25] Wahyudi M and Kristiyanti D A, 2016 Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization J. Theor. Appl. Inf. Technol. 91, 1.

[26] C. W D, 2009 Projects In Computing And Information System A Student’s Guide England: Addison-Wesley.

[27] Vercellis C, 2009 Business Intelligence: Data Mining and Optimization for Decision Making Bus. Intell. Data Min. Optim. Decis. Mak. p. 1–417.

[28] Peraturan Pemerintah no. 41 tentang Pengendalian Pencemaran udara, 1999, Peraturan Pemerintah no. 41 tentang Pengendalian Pencemaran udara, Jakarta.