MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi
(*Journal of Computer Science and Information Technology)*
ISSN : 1978-161X(p); 2477-2550(e)
Volume 14, No. 1(2017), pp 33-40
Received: January 12th 2022; **Accepted**: March 9th 2022; **Avalaible Online**: June 15th 2022

Air Quality Forecasting in DKI Jakarta Using Artificial Neural Network

Asfilia Nova Anggraini, Nisa Kholifatul Ummah, Yessy Fatmasari, Khadijah Fahmi Hayati

***Abstract*— The increase in the use of motorized vehicles increases air pollution conditions, especially in big cities such as the capital city of Indonesia, Jakarta. The pollution that pollutes this city contains various kinds of chemical particles that are harmful to living things when they enter the body. several efforts to reduce this pollution have been carried out, one of which is by identifying the pollutants contained in the air. This study uses data obtained from monitoring stations to predict the content of pollutants in the air at some time in the future. the method used is data mining forecasting with a neural network model. by using rapid miner obtained several graphic descriptions of pollutant conditions in Jakarta that go up and down. pollutant levels of SO_2, CO, PM10 and NO_2 all increased in the November-December period and at the same time period, ozone was at its lowest point. Results from Prediction air quality using Artificial Neural Network with 5 parameters, shown on this pollutant PM10 had an RMSE of 9,477; SO_2 had an RMSE 5,474; CO had an RMSE 8,392; O_3 had an RMSE 18,250; NO_2 had an RMSE 5,171. Can be concluded that the RMSE value of O_3 is higher than the others and the lowest value of NO_2.**

***Index Terms*—Air quality, Air Polutant, Forecasting, Neural Network.**

I. INTRODUCTION

The increase in air pollution caused by vehicles and industrial emissions has been a severe problem, especially for the big city with the higher number of urbanizations, industrial development, and fossil fuel users. This pollution contains several chemical materials that have a severe detriment to human and other living creatures' health, such as sulfur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), and particular matters. The pollutant affects humans, animals, or vegetation and on stone-made material, statues, and monuments[1]. Air pollution is a severe environmental problem in so many parts of this world[2],. A big city in every country of the world has suffered from this in recent years, not excluding Indonesia's capital city Jakarta.

Daerah Khusus Ibukota Jakarta, known as DKI Jakarta, is an Indonesian Capital City located on Java Island. Jakarta is a province which is divided into 5 cities and 1 district of the Kepulauan Seribu. Based on the 2020 population census, the population reaches 10.7 million with an area of ± 661.52 km^2. Jakarta is currently the center of all politics, economy, and some important aspects in Indonesia. The higher population number causes the number of vehicles to be directly proportional so that air pollution concentration increases. The road must be filled with numerous cars, motorbikes, and diesel fuel-use vehicles. The high population can also be identified as an increasing industrial environment as Jakarta is a big city. Many people crave to make a living by earning money from this city, making the villagers move. The industry sees this as their opportunity to build the factory and make air pollution worst as they do not treat the waste properly[3].

Based on air quality control site iqair.com, based on data obtained in 2019, from April to December it can be categorized as the worst air quality with May, June, July, September, October entering the unhealthy category on the PM2.5 parameter, which is 55.5 µg/m^3. This year, the first three months of air quality was the cleanest, although still high with 24.2 µg/m^3 the lowest average in January. Jakarta's air quality reached level 182 Air Quality Index US, categorized as unhealthy air quality. In recent years, as air pollution is a serious problem, some researchers have made air quality research efforts. Mostly about forecasting the air pollution with some methods. For example, presented deep air quality forecasting using deep hybrid

Manuscript received September 03, 2021. This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University.

Asfilia Nova Anggraini is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University, Malang, Indonesia (email 18650040@uin-malang.ac.id)

Nisa Kholifatul Ummah, was with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University, Malang, Indonesia (email 18650065@uin-malang.ac.id)

Yessy Fatmasari, was with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University, Malang, Indonesia (email 18650085@uin-malang.ac.id)

Khadijah Fahmi Hayati Holle is the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University, Malang, Indonesia (email khadijah.holle@gmail.com)

learning[4]. They provide ideas for air quality forecasting problems called the Deep Air Quality Forecasting Framework (DAQFF), which examines multivariate air quality data, such as dynamic, spatial-temporal, and nonlinear characteristics in deep learning models.

This research proposes air quality forecasting using an Artificial Neural Network to predict the air quality in years to come. Artificial Neural Network (ANN) is one of the artificial representations of the human brain that always tries to simulate the learning process in the human brain[5]. In this study, we used a dataset from Jakarta open data. This data was taken every day in 2012 until 2020 from five different air observer stations in DKI Jakarta. This dataset contains several pollutants with their index, worst index point in the critical side, and air condition in that day data taken.

II. DATASET

The data that is implemented in the air quality prediction research using an artificial neural network is obtained from the Jakarta open data available on the https://data.jakarta.go.id/. The data was taken from 5 Air Quality Monitoring Stations (SPKU) based on the Air Pollutant Standard Index (ISPU) under the supervision of the DKI Jakarta Provincial Environmental Service. The data contains 5 parameters that can be used as indicators in this study, among others, PM10 (Particulate), CO (Carbon Monoxide), SO_2 (Sulfide in the form of SO_2), O_3 (Ozone) and NO_2 (Nitrogen dioxide). In addition, there is also data on the date of air quality measurement, measurement location, the highest value of all parameters and categories of results of the calculation of the air pollution standard index.

Previous research has focused on the recognition of air quality patterns in Malaysia by determining the ability to predict the air pollution index (API). The data source for the study was taken from the Malaysian Department of the Environment (DEO). The study collected data for 6 years, starting (2005-2011), this was done to analyze the main component (PCA). From the results of PCA analysis, showed that the most significant parameters were CH4, NmHC, THC, 03, and PM10. After that, to determine the air pollution index (API) prediction, a combination of PCA and artificial neural networks is needed. It was concluded that from this study, ANN PCA has the ability to predict better in determining the API by using fewer variables so as to produce R2 and Root Mean Square Error (RMSE) values of 0.618 and 10.017[6].

III. RESEARCH METHODOLOGY

This research was conducted in several stages, including:

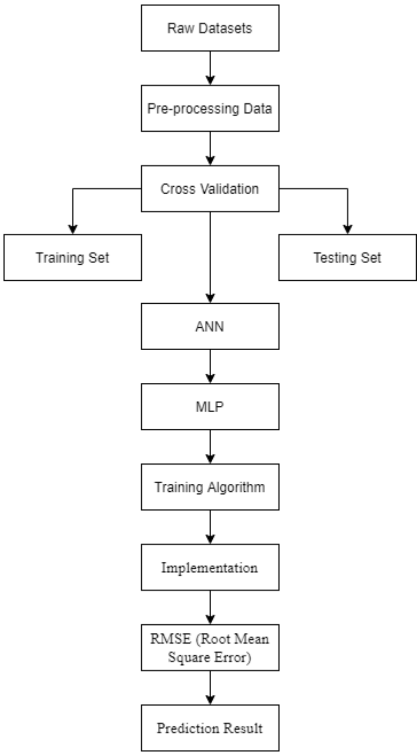

Fig. 1. Conseptual Framework

Figure 1 shows the flow in the implementation process. In the implementation of this research, starting from the raw datasets, pre-processing, prediction process, performing RMSE testing and implementation.

This research data comes from the DKI Jakarta Provincial Environment Agency.

A. *Raw Datasets*

The data collected is air quality data for 5 (five) years, which comes from 5 air quality monitoring stations (SPKU) in DKI Jakarta. The data collected for each SPKU are; PM10 (Particulate), CO (Carbon Monoxide), SO_2 (Sulfide in the form of SO2), O_3 (Ozone) and NO_2 (Nitrogen dioxide).

B. *Pre-Processing Data*

At this stage, the air quality data from 2012-2020 are grouped into 2 types of data, namely training data and testing data as input data. This stage also aims to prepare data so that it can be used in predictions. Next perform normalization or data transformation. The normalization or transformation process has the aim of simplifying calculations and getting more accurate prediction results[7]. Furthermore, the data will be analyzed to be able to measure the effect of more than one independent variable on the dependent variable using multiple linear regression. The formula used in the regression analysis is as follows[8].

$$\mathbf{Y = \alpha + b1X + b2X2 + \ldots + bnXn} \quad (1)$$

C. Prediction Process

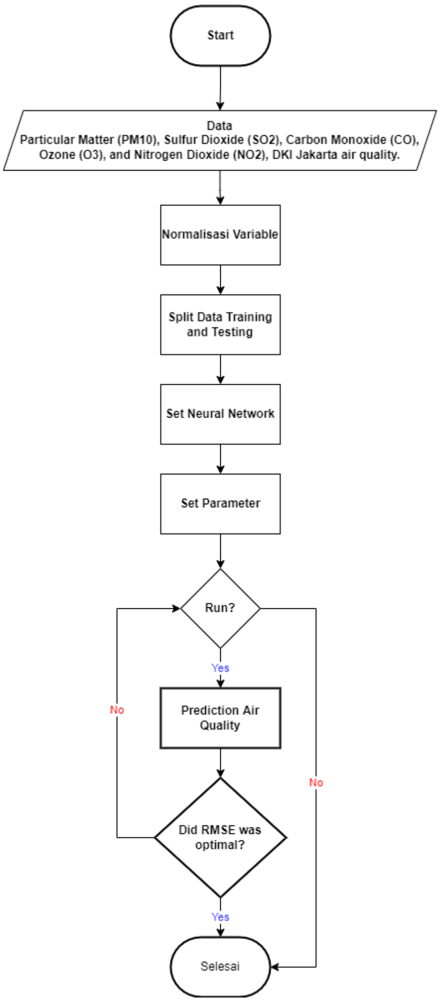

Fig. 2. System Design

Data from Jakarta's open data were entered into the database, then used as input for the ANN model. Schematic representation of an artificial neural network.

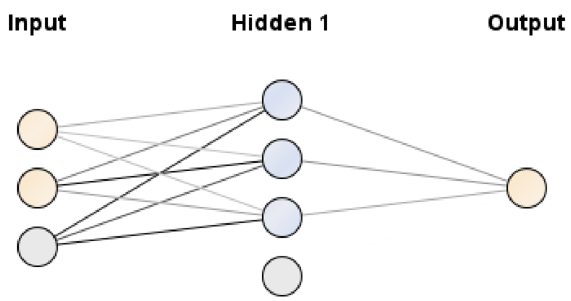

Fig. 3. ANN Design Model

ANN is a statistical and non-statistical data modeling tool. ANN can perform complex relationship modeling (complex) between input and output with the aim of finding patterns in the data[9]. consists of: input layer (independent variable), hidden layer and, output layer (dependent variable).

a. Input Layer, serves to receive input data from outside which will be processed in the next process.
b. Hidden Layer, in this layer the resulting output cannot be observed directly.
c. Output Layer, the output of this layer is the ANN solution to a problem[10].

The ANN model used is feedforward which is used to train data with test data based on parameter validation utilizing the equation:

$$v_m = \sum_{i=1}^{L} w_{im} x_i + b_i \quad (2)$$

After all the results are collected in one block, then it is activated by a linear or nonlinear function that has the following input:

$$y_m = f(v_m) \quad (3)$$

$$f(v_m = \frac{1}{1 + \exp(-\beta v_m)'})` \quad (4)$$

The sigmoid function is a function that is often applied in ANN applications the slope parameter of the function.

Compute the learning errors for every neuron layer by layer:

$$\delta_m = (D_m - y_m) v_m \quad (5)$$

And next step update weight dan bias using function

$$w_{jk}(t+1) = w_{jk} + \eta \delta_k y_j + \alpha[w_{jk}(t) - w_{jk}(t-1)] \quad (6)$$

D. Performing RMSE Testing

At this stage, the prediction results that have been obtained are tested by looking at the error rate in the system. Tests are carried out on training data and testing data. The test aims to determine whether the input, process and output work systems are in accordance with the expected goals. How to calculate the error value is used RMSE (Root Mean Square Error). The smaller the RMSE value, the better the prediction performance.

$$\textbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Xesti - X\ meas)^2}{N}} \quad (7)$$

E. Implementation

At this stage, the results of the ANN prediction research are applied to Rapidminer to make it easier for users to operate the model that has been built.

IV. Result and Discussion

This study uses pollutant data from air pollutants in the form of PM10, SO_2, CO, O_3 and NO_2 from 2012 to 2020 as many as 3290 records of each parameter taken from station 1 DKI Jakarta. Data preprocessing is done by eliminating missing values using the average of all data for one year.

The data that has been prepared at the preprocessing stage will be processed by using rapidminer. Aiming to training and testing the data, we combine all data to get raw data for every year, then it divided again into each kind of pollutant material to get each pollutant prediction. After deviding the data into each kind of pollutant which is univariat data, we changed them into multivariat data using multidimentional method based on each independent data from each pollutan as XT and run as a label, and for prediction process used two dimension, XT-1 and XT-2.

The result is a several number including the past data. To get the predistion result, take from the excess number from the given data before.

The following are the experimental results on each parameter. The experimental results shown are forecasting air pollutant levels in 2021 starting from April to December by including data obtained from January to March. Blue line represent actual data from the data souce and the orange one is the prediction result.

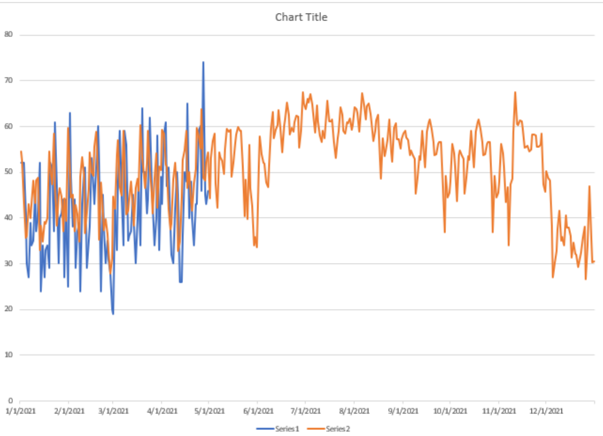

Fig. 4. PM10 Prediction

Figure 4 shows the level of pollutant levels with the code PM10 in 2021. PM10 pollutant levels in the capital city of DKI Jakarta are expected to gradually increase and have a peak around month July. Experiments on this pollutant have an RMSE of 9,477.

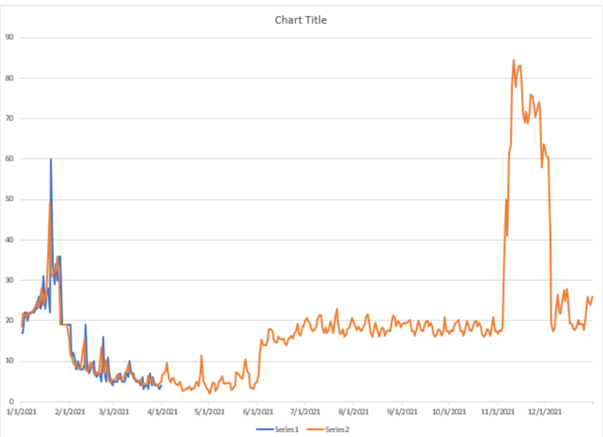

Fig. 5. SO_2 Prediction

Figure 5 shows the level of SO_2 or sulfate pollutant levels in 2021. CO pollutant levels in the capital city of DKI Jakarta are expected to decrease from March-April. In June the pollutant value increased and the value was stable for the next few months, but the highest value spike occurred in November-December with the highest value experienced for a year. Experiments on this pollutant had an RMSE of 5,474.

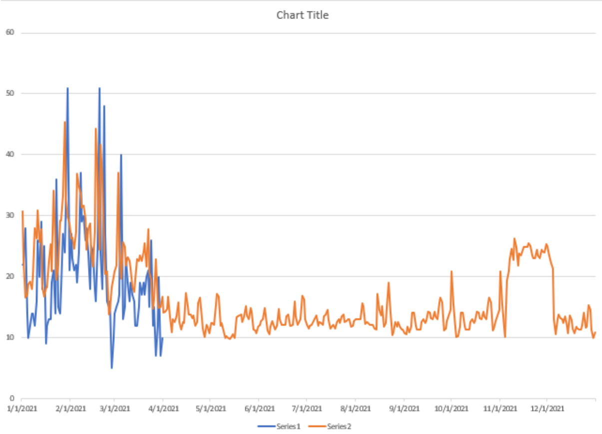

Fig. 6. CO Prediction

Figure 6 shows the level of CO pollutant levels in 2021. CO pollutant levels in the capital city of DKI Jakarta are expected to have the lowest values from April to November and will increase around November-December. Experiments on this pollutant had an RMSE of 8,392.

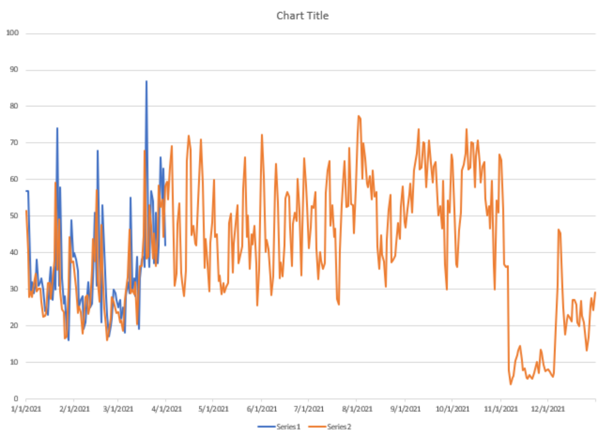

Fig. 7. O3 Prediction

Figure 7 shows the level of O_3 or ozone levels in 2021. Ozone levels in the capital city of DKI Jakarta are expected to be in a certain range during April-November. the displayed value is quite stable and changes vary over time. unlike some previous pollutants which have increased, ozone levels will decrease over the November-December time period. Experiments on this pollutant have RMSE of 18,250.

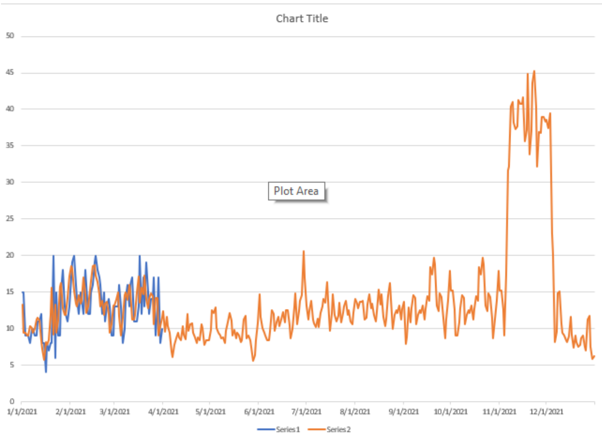

Fig. 8. NO2 Prediction

Figure 8 shows the predicted level of NO_2 levels in 2021. NO_2 levels in the capital city of DKI Jakarta are estimated to be below 25 in the April-November period and will increase and have the highest value in the November-December period. Experiments on this pollutant had an RMSE of 5,171.

The RMSE value in this experiment varies because the experiment is carried out on each pollutant so that it produces different values.

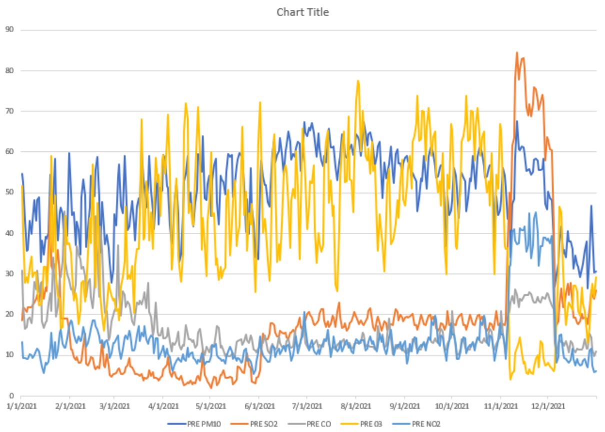

Fig. 9. Polutant Prediction

Figure 9 shows a graph of the predicted levels of several pollutants in the air in 2021. In the figure it can be seen that the PM10 pollutant prediction is the only one that has increased in the April-November time period. while the predicted value of ozone appears to be more stable with a range of values that do not change over that time period. the other three politans actually tended to experience a decline in the April-June period and slightly increased around the months of June-early November. All pollutants except ozone experienced a fairly high value increase in the late November-early December period. This change looks very striking in CO, SO_2 and NO_2 pollutants. In this period of time ozone actually experienced its lowest value in one year.

Table 1 shows that the RMSE value of O_3 is higher than the others and the lowest value of NO_2.

Table 1. RMSE value

No.	Polutan	RMSE
1.	PM10	9,477
2.	SO2	5,474
3.	CO	8,392
4.	O_3	18,250
5.	NO_2	5,171

V. Conclusion

The following are the conclusions of the air pollution forecasting experiment using the ANN method. The experimental results of each pollutant have a different RMSE but all values are below 20. The pollutant element PM10 was the only one that experienced a constant increase during the April-November 2021 period. The three pollutants, namely SO_2, CO and NO_2, decreased during April-June and increased thereafter. All pollutant values increased around the time period November-December except ozone which actually decreased.

References

[1] F. T. Sipil, U. Muhammadiyah, S. Utara, J. Kapten, M. Basri, and G. D. Ii, "PENCEMARAN UDARA AKIBAT KINERJA LALU-LINTAS Air Pollutions Due to Traffic Performance of Motor Vehicles in Medan City," *J. Pemukim.*, vol. 13, no. 1, pp. 13–20, 2018.

[2] F. Ruhiat, D. Heryadi, and A. -, "Strategi NGO Lingkungan Dalam Menangani Polusi Udara di Jakarta (Greenpeace Indonesia)," *Andalas J. Int. Stud.*, vol. 8, no. 1, p. 16, 2019, doi: 10.25077/ajis.8.1.16-30.2019.

[3] F. Jansen, S. L. Sengkey, and Steenie Wallah, "Tingkat Pencemaran Udara CO Akibat Lalu Lintas Dengan Model Prediksi Polusi Udara Skala Mikro," *J. Ilm. MEDIA Eng.*, vol. 1, no. 2, pp. 119–126, 2011.

[4] S. Du, T. Li, Y. Yang, and S. J. Horng, "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework," *arXiv*, 2018, doi: 10.1109/tkde.2019.2954510.

[5] A. Wanto, A. P. Windarto, D. Hartama, and I. Parlina, "Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density," *IJISTECH (International J. Inf. Syst. Technol.*, vol. 1, no. 1, p. 43, 2017, doi: 10.30645/ijistech.v1i1.6.

[6] A. Azid *et al.*, "Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: A case study in Malaysia," *Water. Air. Soil Pollut.*, vol. 225, no. 8, 2014, doi: 10.1007/s11270-014-2063-1.

[7] H. Putra and N. Ulfa Walmi, "Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 100–107, 2020, doi: 10.25077/teknosi.v6i2.2020.100-107.

[8] J. J. Hopfield, *Artificial Neural Networks*, vol. 4, no. 5. 1988.

[9] M. Y. Habibi and E. Riksakomara, "Peramalan Harga Garam Konsumsi Menggunakan Artificial Neural Network Feedforward-Backpropagation (Studi Kasus : PT. Garam Mas, Rembang, Jawa Tengah)," *J. Tek. ITS*, vol. 6, no. 2, 2017, doi: 10.12962/j23373539.v6i2.23200.

[10] Y. Umaidah, "Penerapan Algoritma Artificial Neural Network Dalam Prediksi Harga Saham Lq45 Pt. Bank Rakyat Indonesia, Tbk," *J. Gerbang*, vol. 8, no. 1, pp. 57–64, 2018.