Prediction of air quality in Jakarta during the COVID-19 outbreak using long short-term memory machine learning

Prediction of air quality in Jakarta during the COVID-19 outbreak using long short-term memory machine learning

Wihayati[1] **and F W Wibowo**[2]

[1] Astrophysics and Mathematics Group, Sekip Utara Bulaksumur, Yogyakarta 55281, Indonesia
[2] Informatics Department, Universitas Amikom Yogyakarta, Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55283, Indonesia

Email: wihayati@gmail.com[1], ferry.w@amikom.ac.id[2]

Abstract. Air pollution is one of the world's problems, not just one location. This air pollution is caused by pollutants that are harmful to human health and the environment. Some pollutants are most influential, namely particulate matter, ground-level ozone, carbon monoxide, sulfur dioxide, and nitrogen dioxide. Several countries decided to lock down when the COVID-19 outbreak was announced simultaneously throughout the world like a pandemic. In Jakarta, Indonesia applies large-scale social restrictions (PSBB). The resulting impact is a drastic reduction in air pollution on air quality. This paper aims to predict air quality during the COVID-19 outbreak in Jakarta using long short-term memory (LSTM) machine learning. The evaluation of the LSTM model used in this paper is the root mean square error (RMSE). The results obtained show that the Adam optimizer can bring the prediction results closer to the dataset used.

1. Introduction

Air pollution is an unresolved world problem and will always exist. It is a form of pollution that refers to air contamination causing physical, biological, and chemical changes to the atmosphere [1]. Measurement of pollution levels in an area can be measured with a sensor [2,3]. The sensor is then processed to be sent via the internet of things technology to determine the levels of pollution of particles and gases produced by certain pollutants [4]. The resulting pollution can also be displayed on a map that can be accessed via a browser [5]. Not only that, to analyze the effect of the number of vehicles on the level of pollution produced, it can detect images or videos related to the number of vehicles at a place in traffic lights [6]. So that the setting of traffic lights can be done automatically according to the busiest lane for vehicles to run first. This can regulate the level of pollution produced by the vehicle. Sources of air pollution can be classified into stationary sources and mobile sources. The stationary sources consist of industry, power plants, and households.

Meanwhile, the mobile sources are motor vehicle activities and sea transportation. There are two main types of air pollutants, which are gaseous compounds and solid compounds. It triggers air pollution in industrial areas and big cities, which impacts health including cancer of various organs of the body, heart disease, hypertension, reproductive disorders, respiratory disorders, and so on [7]. There are five most influential pollutants released by United States Environmental Protection Agency (US EPA), namely particle pollution or known as particulate matter, including $PM_{2.5}$ and PM_{10}, ground-level ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), and Nitrogen Dioxide (NO_2).

This paper uses data obtained from the Indonesian government, which only uses PM_{10}, O_3, CO, SO_2, and NO_2.

These pollutants that pollute the air were reported to have had a very drastic decrease during the COVID-19 outbreak [8]. The COVID-19 outbreak has almost affected human life and the environment around the world, including Indonesia. Even though the cure rate for COVID-19 in Indonesia is relatively high than death [9], if it is not watched out for and the spread of it is anticipated, the increase is relatively large in addition to the limitations of hospitals that accommodate sufferers. The predicted increase in COVID-19 sufferers in Indonesia will continue to increase over time if there are no proper regulations. One of the ways to prevent the widespread of COVID-19 in several countries is a lockdown. In Indonesia, the way to anticipate its spread is through large-scale social restrictions (PSBB). Jakarta government did this. Jakarta is one of the big cities in Indonesia, which certainly has an immense contribution to pollution. Various methods were used to solve congestion and to enforce even and odd number plate of vehicle regulations. The existence of this PSBB has reduced the level of pollution in Jakarta. This paper aims to predict the air pollution during the COVID-19 outbreak in Jakarta using machine learning Long Short-Term Memory (LSTM). The model obtained will be evaluated using the root mean square error (RMSE).

2. Related Works

Air quality prediction is a complex issue. Air quality in an industrial area is a challenge for policymakers in making decisions. Of course, these decision-makers need the right information to predict it. The development of machine learning has a place in determining future information on a case, including air quality [10]. Conditions for normal daily activities can be used to predict and monitor air quality in urban areas. One of the methods used to model hourly predictions can use support vector regression (SVR), as was done by [11]. The selection of the right dataset and variables needs to be considered to model an accurate air pollution prediction. There are two main classes that differ in the perspective of air pollutant concentrations, namely estimation, and forecasting. The ensemble learning algorithm and linear regression are suitable for pollution estimation, while the neural network (NN) and support vector machines (SVM) are suitable for forecasting [12]. Air pollution prediction systems PM_{10} and $PM_{2.5}$ have been carried out by [13] to predict Korea's occurrence. The method used in this prediction is deep learning using stacked autoencoders for learning and training data. It proves that the predictions used in air quality modeling using machine learning are quite acceptable.

To suppress the spike in the COVID-19 outbreak, various government regulations in the world have been taken. This is quite effective in reducing the rate of people traveling, and economic activity has also decreased. But on the other hand, the air quality during COVID-19 had a pretty good impact, marked by a reduction in air pollution during the months that these prohibitions were implemented [14]. The study conducted by [15] used the LSTM method to predict air quality in Madrid. Each pollutant has a different behavior, so it is necessary to pay attention to the machine learning implementation. Each type of pollutant behaves differently at each location. It is different from the research in this paper even though the method used is the same, namely LSTM. The research in this paper takes the air quality in Jakarta at the time of COVID-19, which uses data from 6 locations. The rate of death cases caused by COVID-19 is predicted to increase over time [16], so the necessary mitigation must be taken to anticipate it.

3. Research Methodology

The principles used in this research are preprocessing data, initializing parameters, training LSTM networks, and testing the test data. The root mean square error (RMSE) was used to determine the error value between the model and the original data. This paper compares three optimizers, consisting of adaptive moment estimation (Adam), stochastic gradient descent (SGD), and root mean square propagation (rmsprop), to get the RMSE and epoch values for 10, 20, 30, 50, and 100, respectively.

3.1. Preprocessing dataset

The dataset used in this paper is obtained from the website https://data.jakarta.go.id/dataset/indeks-standar-pencemaran-udara-ispu-tahun-2020 [17]. This dataset consists of the air pollution standard index (*indeks standar pencemar udara* (ISPU)) gotten from the six locations of the air quality monitoring in Jakarta, which are Bunderan HI, Kelapa Gading, Jagakarsa, Lubang Buaya, Kebun Jeruk, and DKI Jakarta. The dataset used in this paper is taken from January 2020 until early June 2020, which is the COVID-19 outbreak has emerged. Some of the data contained in the dataset is missing. Because the missing data has no significant effect, the disappeared datasets's rows are deleted to obtain continuous data. Likewise, for double data, the data selected is the first data only. Information on the dataset of air quality at the Bunderan HI can be seen in Table 1.

Table 1. Air quality dataset information of PM_{10}, SO_2, CO, O_3, and NO_2 in Bunderan HI

	PM_{10}	SO_2	CO	O_3	NO_2
Number of data	148	148	148	148	148
Mean	35.196	11.541	20.047	34.993	19.351
Standard deviation	18.518	9.244	17.845	18.881	18.337
Minimal value	3	1	3	8	4
Maximal value	74	60	87	90	87

Table 1 shows the amount of data for each pollutant. The amount of data used is 148 data. In contrast, the information on the air quality dataset in Kelapa Gading is shown in Table 2.

Table 2. Air quality dataset information of PM_{10}, SO_2, CO, O_3, and NO_2 in Kelapa Gading

	PM_{10}	SO_2	CO	O_3	NO_2
Number of data	147	147	147	147	147
Mean	40.463	13.143	23.905	43.163	19.306
Standard deviation	19.521	7.387	22.956	26.094	24.792
Minimal value	7	1	4	2	2
Maximal value	86	42	124	124	124

Table 2 shows the amount of data used in Kelapa Gading for observation amounted to 147 data. At the same time, the information on the air quality dataset in Jagakarsa is shown in Table 3.

Table 3. Air quality dataset information of PM_{10}, SO_2, CO, O_3, and NO_2 in Jagakarsa

	PM_{10}	SO_2	CO	O_3	NO_2
Number of data	133	133	133	133	133
Mean	42.368	18.271	35.714	48.744	20.398
Standard deviation	16.190	6.139	23.944	26.870	28.907
Minimal value	15	10	12	4	2
Maximal value	76	51	137	126	137

Table 3 shows the amount of data used in Jagakarsa for observation totaling 133 data. In comparison, the information on the air quality dataset in Lubang Buaya is shown in Table 4.

Table 4. Air quality dataset information of PM_{10}, SO_2, CO, O_3, and NO_2 in Lubang Buaya

	PM_{10}	SO_2	CO	O_3	NO_2
Number of data	122	122	122	122	122
Mean	51.975	25.607	31.648	54.492	20.508
Standard deviation	19.364	6.534	34.097	31.554	38.018
Minimal value	6	6	8	2	2
Maximal value	111	40	197	191	197

Table 4 shows the air quality information used at Lubang Buaya is 122 data. Meanwhile, the dataset information used in Kebun Jeruk is shown in Table 5.

Table 5. Air quality dataset information of PM_{10}, SO_2, CO, O_3, and NO_2 in Kebun Jeruk

	PM_{10}	SO_2	CO	O_3	NO_2
Number of data	141	141	141	141	141
Mean	42.582	16.837	26.674	65.922	17.220
Standard deviation	16.235	11.447	25.749	32.130	27.393
Minimal value	5	5	5	3	2
Maximal value	76	63	179	182	179

Table 5 shows the data used for observation in Kebun Jeruk is 141 data. Whereas the air quality dataset information in the Province of DKI Jakarta is shown in Table 6.

Table 6. Air quality dataset information of PM_{10}, SO_2, CO, O_3, and NO_2 in Province of DKI Jakarta

	PM_{10}	SO_2	CO	O_3	NO_2
Number of data	168	168	168	168	168
Mean	56.357	28.613	31.250	75.065	18.845
Standard deviation	18.271	8.753	22.625	34.753	23.550
Minimal value	9	10	12	3	5
Maximal value	111	63	179	191	179

The air quality data number used in DKI Jakarta is 168 data, the largest number of the others. The air quality levels can be seen in Table 7.

Table 7. Air quality levels [18,19]

Values of Index	**Air Pollution Level**		**Implications**	**Cautionary Statement (for PM_{10})**
	US EPA	**Indonesia**		
0-50	Good	Good	Air quality is satisfactory and no risk	There is no effect
51-100	Moderate	Moderate	Air quality is acceptable ad may be a risk for some people	There is a decrease in visibility
101-150	Unhealthy for sensitive groups	Unhealthy	Some people may have health effects, and the general public is less affected	Visibility drops, and there is dust fouling everywhere
151-200	Unhealthy		Some people may experience more severe health effects	
201-300	Very unhealthy	Very unhealthy	Health effects are increased for everyone	Increased sensitivity in patients with asthma and bronchitis
300+	Hazardous	Hazardous	Health warning for everyone	Dangerous levels for all exposed populations

Table 7 shows that the higher the air quality index value, the greater air pollution level and the greater the health risk for the people and environments. There are differences in the standard levels used between the United States Environmental Protection Agency (US EPA) and Indonesia. However, they are almost the same, namely the index value between 101-200. At that level, the US EPA divides it into two air pollution levels, namely unhelpful for sensitive groups and unhealthy, while in Indonesia, there is only one level, namely unhealthy.

3.2. Long short-term memory network machine learning

Long short-term memory (LSTM) is a type of processing model from the evolution of recurrent neural network (RNN) created by Hochreiter and Schmidhuber in 1997. This model modifies the RNN by adding a memory unit used to store information for a long time. This model is proposed as a solution to the vanishing gradient in RNN when processing long sequence data. Due to the vanishing gradient problem, the RNN can't capture long-term dependencies, thus reducing the accuracy of using predictions on the RNN. LSTM overcomes these problems because it can manage the memory in each input using memory units and gate units, as shown in figure 1.

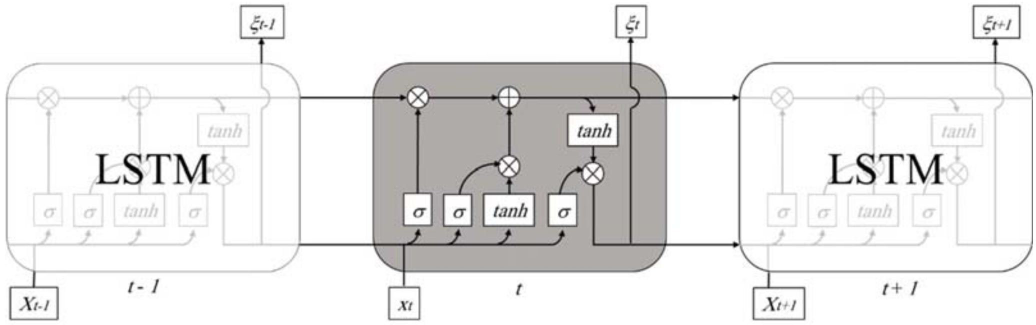


Figure 1. Long short-term memory model

The LSTM module has four gates: input gates, cell gates, forget gates, and output gates. The memory cell is designed using linear and logistic units with multiplicative interactions. As in the RNN, this LSTM network consists of modules with repetitive processing formed from the LSTM. The computational working principle of LSTM can be stated as follows,

- The value of the input can only be stored in the memory cell if the gate input allows. The activation function in forget gates is a sigmoid activation function that has outputs 0 and 1. If the output is 0, all data will be discarded, whereas if the output is 1, all data will be stored. Calculation of the input gate α_t and candidate values ς_t of the cell state is done using equations 1 and 2, respectively.

$$\alpha_t = \sigma\left(\omega_\alpha x_t + \upsilon_\alpha \xi_{t-1} + \beta_\alpha\right) \quad (1)$$

Where σ is a sigmoid function, ω_α is a weight of input at time t, x_t is an input value at time t, υ_α is a weight of output at time t, ξ_{t-1} is an output value at time t-1, and β_α is a bias in the gate input.

$$\varsigma_t = \tanh\left(\omega_\varsigma x_t + \upsilon_\alpha \xi_{\varsigma-1} + \beta_\varsigma\right) \quad (2)$$

Where tanh is a hyperbolic tangent function, ω_ς is a weight of input at the cell ς, υ_α is a weight of output value from cell to ς-1, $\xi_{\varsigma-1}$ is an output value from cell to ς-1, and β_ς is a bias in the cell to ς.

- The forget gate value ψ_t is calculated as shown in equation 3.

$$\psi_t = \sigma\left(\omega_\psi x_t + \upsilon_\psi \xi_{t-1} + \beta_\psi\right) \quad (3)$$

Where ω_ψ is a weight of input at time t, υ_ψ is a weight of output at time t-1, ξ_{t-1} is an output value at time t-1, and β_ψ is a bias in the gate input.

- The memory cell state ζ_t could be calculated using equation 4.

$$\zeta_t = \alpha_t \cdot \varsigma_t + \psi_t \cdot \zeta_{t-1} \quad (4)$$

Where ζ_{t-1} is a memory cell state at the previous cell.

- After the new memory cell state has been resulted, the output gate φ_t could be calculated as shown in equation 5.

$$\varphi_t = \sigma\left(\omega_\varphi x_t + \upsilon_\varphi \xi_{t-1} + \beta_\varphi\right) \tag{5}$$

Where ω_φ is a weight of output at time t, υ_φ is a weight of output at time $t-1$, ξ_{t-1} is an output value at time $t-1$, and β_φ is a bias in the gate output.

- The final output ξ_t can be calculated as equation 6.

$$\xi_t = \varphi_t \tanh \zeta_t \tag{6}$$

3.3. Optimizer

The adaptive moment estimation (Adam) optimization algorithm is an extension to stochastic gradient descent (SGD) which is used also in deep learning. It has been introduced by Kingma and Ba [20]. The Adam is an optimization algorithm that develops by leveraging the advantages of the adaptive gradient (AdaGrad) and root mean square propagation (RMSProp) algorithms. Learning parameters based on the first mean as in RMSProp, Adam also uses the second mean of the gradient or uncentered variance. This algorithm calculates the exponential moving average of the gradient and its quadratic gradient and the parameters γ_1 and γ_2 control the moving average decay rate. The pseudo code of Adam can be written as below,

τ	# learning rate
$\gamma_1, \gamma_2 \in [0,1)$	# exponential decay values for the moment estimates
Φ_0	# gradient parameter
$\rho(\phi)$	# stochastic function with parameter ϕ
ϕ_0	# initialize vector parameter
κ	# tolerance parameter $\kappa > 0$ for numerical stability
$\mu_0 \leftarrow 0$	# initialize first vector moment estimate
$\eta_0 \leftarrow 0$	# initialize second vector moment estimate
$t \leftarrow 0$	# time-step to zero
while ϕ_t not converged do	
$t \leftarrow t+1$	# update time-step
$\Phi_t \leftarrow \nabla_\phi \rho_t(\phi_{t-1})$	# compute gradient of objective at time-step t
$\mu_t \leftarrow \gamma_1 \mu_{t-1} + (1-\gamma_1)\Phi_t$	# update the first moment estimate
$\eta_t \leftarrow \gamma_1 \eta_{t-1} + (1-\gamma_1)\Phi_t^2$	# update the second moment estimate
$\overline{\mu_t} \leftarrow \dfrac{\mu_t}{1-\gamma_1^t}$	# create unbiased estimate $\overline{\mu_t}$
$\overline{\eta_t} \leftarrow \dfrac{\eta_t}{1-\gamma_2^t}$	# create unbiased estimate $\overline{\eta_t}$
$\phi_t \leftarrow \dfrac{\phi_{t-1} - \tau\overline{\mu_t}}{\sqrt{\overline{\eta_t}} + \kappa}$	# update objective parameters
end while	
return ϕ_t	# resulting parameters

3.4. Root Mean Square Error (RMSE)

The system performance of the error rate of the prediction results is calculated using the root mean square error (RMSE). In this calculation, the best model of prediction accuracy is the model with the smallest RMSE value or close to 0. This equation is shown in equation 7.

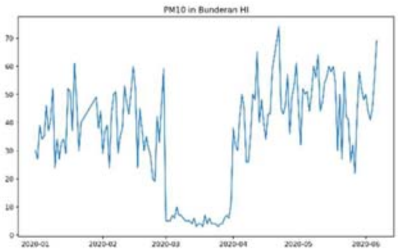
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\Omega_i - \Lambda_i)^2}{\mathrm{N}}} \tag{7}$$

IOP Conf. Series: Earth and Environmental Science **704** (2021) 012046 doi:10.1088/1755-1315/704/1/012046

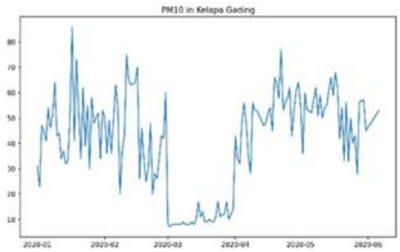
where Ω_i is the value of the predicted model, Λ_i is the value of the observed i^{th} data point, and N is the data amount.

4. Result and Discussion

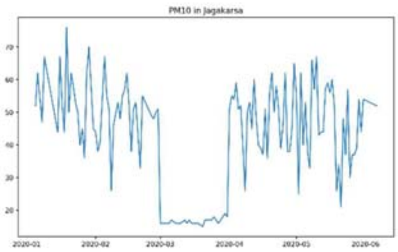
The data used in this paper only uses PM_{10} and O_3 data from six locations in Jakarta. The graphs of the PM_{10} data for these six locations are shown in Figure 2.



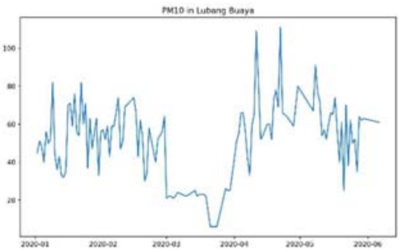
(a)



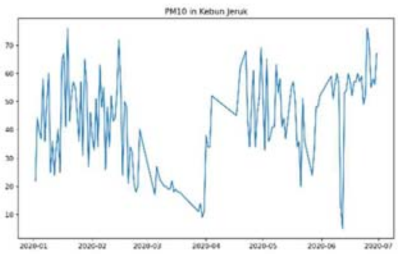
(b)



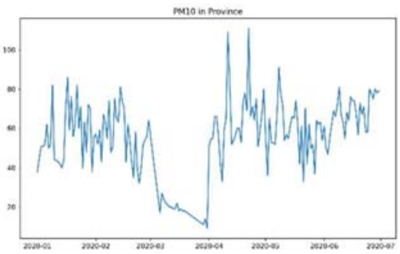
(c)



(d)



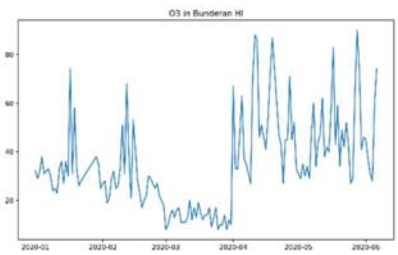
(e)



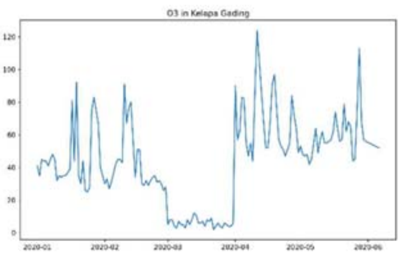
(f)

Figure 2. Air quality of PM_{10} in Jakarta: (a) Bunderan HI, (b) Kelapa Gading, (c) Jagakarsa, (d) Lubang Buaya, (e) Kebun Jeruk, and (f) Province of DKI Jakarta.

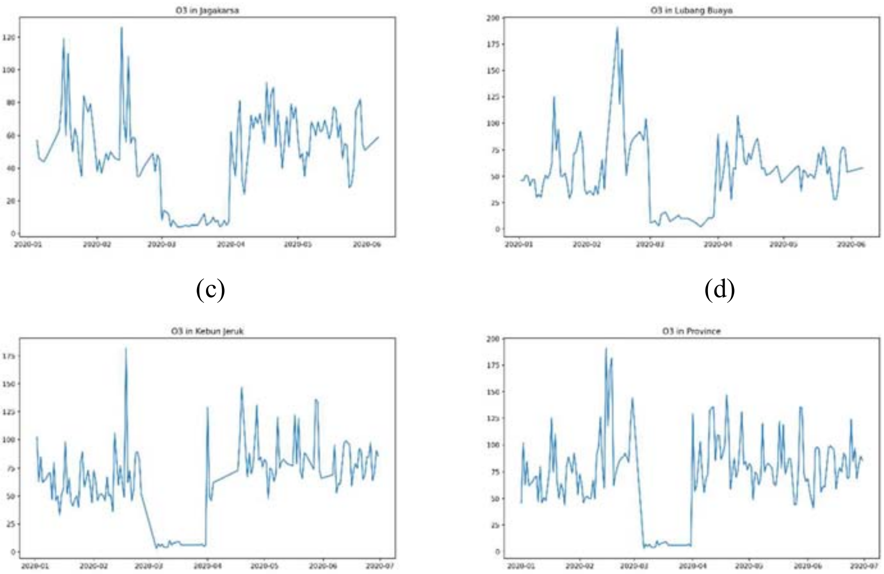
While the graphs of the O_3 data are shown in Figure 3.



(a)



(b)



(c) (d)

(e) (f)

Figure 3. Air quality of O_3 in Jakarta: (a) Bunderan HI, (b) Kelapa Gading, (c) Jagakarsa, (d) Lubang Buaya, (e) Kebun Jeruk, and (f) Province of DKI Jakarta.

From Figures 2 and 3, it can be seen that in the period March-April, there was a drastic decrease in PM_{10} and O_3 pollution. It is because, at that time, a large-scale social restriction (PSBB) was implemented in Jakarta, which took effect Friday (10/4/2020) until Thursday (23/4/2020) [21]. It was done after the global COVID-19 pandemic announcement by the world health organization (WHO) and patients were found and affected in Jakarta. However, after the PSBB period ended, the pollutions of PM10 and O3 increased just like before.

The model testing in this paper analyzes the impact of the parameters on the accuracy obtained. The parameters tested were the composition of the data, the number of time series patterns, the number of hidden neurons, and the amount of epoch to determine the LSTM weight that has the lowest RMSE. The parameter values tested were 70% training data and 30% test data. The number of hidden neurons in the LSTM is 4, the magnitude of the epoch used is 10, 20, 30, 50, and 100, respectively. The results of the RMSE obtained for the PM_{10} pollutant can be seen in Table 8.

Table 8. RMSE results of LSTM using various optimizer and epoch of air quality in Jakarta from PM_{10} data

Location	Epoch	RMSE of Train Score (PM_{10})			RMSE of Test Score (PM_{10})		
		Adam	SGD	rmsprop	Adam	SGD	rmsprop
Bunderan HI	10	15.06	15.56	12.12	17.21	17.88	13.91
	20	11.70	13.82	11.44	13.70	15.62	12.76
	30	11.49	15.23	11.36	12.45	18.71	12.70
	50	10.70	11.38	11.37	12.34	14.05	12.43
	100	**9.82**	11.02	11.43	**12.14**	12.48	12.67
Kelapa Gading	10	16.56	18.56	15.44	15.43	21.04	15.04
	20	14.70	16.08	14.61	12.85	15.84	12.92
	30	14.70	18.74	14.32	12.52	20.27	12.08
	50	13.89	15.60	14.55	11.56	14.99	11.80
	100	**13.73**	14.02	13.88	**11.45**	11.76	11.68

Jagakarsa	10	12.88	13.77	13.41	12.86	13.01	12.58
	20	11.75	14.29	11.80	13.77	12.87	13.52
	30	11.52	12.41	11.55	14.06	13.08	14.41
	50	10.98	11.15	11.51	13.86	13.37	14.54
	100	**10.75**	10.87	11.08	**13.60**	13.67	14.18
Lubang Buaya	10	15.22	16.78	14.49	17.62	20.61	17.61
	20	14.89	17.11	14.45	18.34	21.43	17.35
	30	14.13	17.62	14.09	16.71	21.87	16.43
	50	14.05	15.98	13.95	17.17	18.97	17.01
	100	**13.79**	14.96	13.84	**16.74**	18.13	16.77
Kebun Jeruk	10	14.44	16.08	14.50	13.62	16.29	13.61
	20	14.44	16.17	14.81	13.57	15.91	14.29
	30	14.31	16.23	15.24	13.01	16.05	15.03
	50	14.27	15.51	14.38	13.34	15.56	13.15
	100	14.25	14.94	**14.15**	13.29	14.69	**12.87**
Province of DKI Jakarta	10	17.60	18.12	16.12	13.01	14.25	12.91
	20	15.64	20.16	15.95	11.92	17.07	12.61
	30	15.31	17.75	15.55	12.14	14.18	12.71
	50	15.07	15.73	15.26	12.20	12.21	11.93
	100	**14.78**	16.40	15.32	**11.84**	13.02	12.43

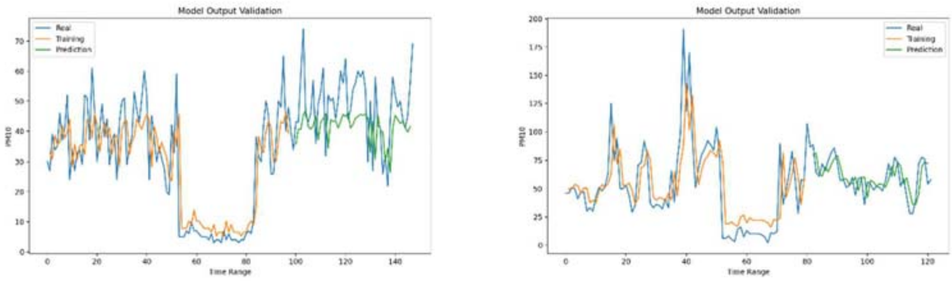
Meanwhile, the results of the RMSE obtained for the O_3 pollutant can be seen in Table 9.

Table 9. RMSE results of LSTM using various optimizer and epoch of air quality in Jakarta from O_3 data

Location	Epoch	RMSE of Train Score (O_3)			RMSE of Test Score (O_3)		
		Adam	SGD	rmsprop	Adam	SGD	rmsprop
Bunderan HI	10	14.05	17.86	14.84	19.04	26.28	21.06
	20	14.02	16.22	13.90	19.15	22.11	19.53
	30	13.28	14.73	13.37	17.09	21.50	18.02
	50	13.20	16.08	13.29	16.92	23.98	17.33
	100	**13.20**	13.55	13.27	**16.31**	18.32	16.93
Kelapa Gading	10	17.30	23.50	16.49	18.43	31.51	16.09
	20	16.63	23.43	17.16	16.18	30.32	17.67
	30	16.15	21.52	16.50	14.87	28.24	15.50
	50	15.99	18.67	16.25	14.65	21.72	14.87
	100	**15.63**	16.05	16.23	**14.61**	15.32	14.66
Jagakarsa	10	27.29	29.29	22.46	19.19	21.40	16.57
	20	20.69	22.46	21.25	13.80	16.08	14.66
	30	19.71	24.67	20.61	13.58	17.21	14.42
	50	20.50	21.47	20.02	13.85	13.93	13.65
	100	**18.91**	20.55	19.61	**12.96**	13.99	13.45
Lubang Buaya	10	28.21	36.03	28.92	12.08	16.94	13.15
	20	26.26	35.23	28.41	**11.43**	17.76	13.34
	30	27.80	34.93	26.63	11.98	15.20	11.94
	50	25.34	29.91	25.72	11.52	13.72	11.61
	100	**25.19**	26.66	25.21	11.53	12.95	11.58
Kebun Jeruk	10	29.63	31.84	29.89	26.21	32.25	28.01
	20	28.80	32.76	28.84	24.56	31.27	25.81
	30	28.31	31.50	28.19	24.19	30.87	23.95
	50	28.10	30.94	28.21	23.69	32.52	22.71
	100	**26.48**	29.10	27.44	**21.75**	23.13	23.96

Province of DKI	10	35.96	37.85	32.01	21.07	22.87	20.82
Jakarta	20	31.22	36.43	31.09	21.68	22.13	21.59
	30	31.11	36.16	31.16	21.96	22.45	22.24
	50	30.64	33.13	30.87	21.70	20.89	22.56
	100	30.09	30.78	**29.88**	**21.65**	21.89	21.87

The validation of the model output with the best RMSE value for each data from PM_{10} and O_3 can be shown in Figure 4.



(a) (b)

Figure 4. Prediction model using LSTM: (a) Bunderan HI for PM_{10} and (b) Lubang Buaya for O_3.

5. Conclusion

This paper has produced a prediction model for air quality in Jakarta during the COVID-19 pandemic using machine learning LSTM networks. The best prediction model is obtained with Adam's optimizer. The RMSE value generated from the Adam and rmsprop optimizer has almost the same value but is superior to Adam.

6. References

[1] Ramondt S and Ramírez A S 2020 Media reporting on air pollution: health risk and precautionary measures in national and regional newspapers *Int. J. Environ. Res. Public Health* 17 6516 doi: 10.3390/ijerph17186516

[2] Wibowo F W 2014 Pollution Instrumentation Using Global Positioning System and Data Logger Based-On Propeller *Advanced Science Letters* vol 20 no 10-20 pp 2082-2086 doi: 10.1166/asl.2014.5629

[3] Wibowo F W and Purwacandra P P 2014 Carbon Monoxide Pollution Detection and Measurement Using Knowledge-Based and Probability Approaches *Proceedings of 7th International Seminar on Industrial Engineering and Management* (*7th ISIEM*)

[4] Wibowo F W and Hidayat F 2017 A low-cost home automation system based-on internet of things *Journal of Telecommunication, Electronic and Computer Engineering* vol 9 issue 2-4 p 155-159

[5] Aditya T W and Wibowo F W 2016 Wireless sensor network and geographic information system based monitoring system *Asian Journal of Information Technology* vol 15 issue 15 p 2571-2577 doi: 10.3923/ajit.2016.2571.2577

[6] Wibowo F W and Purwacandra P P 2015 Object tracking using initial data to count object image based-on wireless sensor network *Advanced Science Letters* vol 21 issue 1 p 112-116 doi: 10.1166/asl.2015.5742

[7] Rajper S A, Ullah S, and Li Z 2018 Exposure to air pollution and self-reported effects on Chinese students: A case study of 13 megacities *PLOS ONE* 13(3) e0194364 doi: 10.1371/journal.pone.0194364

[8] Ching J and Kajino M 2020 Rethinking air quality and climate change after COVID-19 *Int J Environ Res Public Health* 17(14) p 5167 doi: 10.3390/ijerph17145167

[9] Wibowo F W and Wihayati 2021 Prediction modelling of COVID-19 outbreak in Indonesia using a logistic regression model *J. Phys.: Conf. Ser.* 1803 012015

[10] Castelli M, Clemente F M, Popovič A, Silva S, and Vanneschi L 2020 A Machine Learning Approach to Predict Air Quality in California *Complexity* vol 2020 Article ID 8049504 23 pages doi: 10.1155/2020/8049504
[11] Mahalingam U, Elangovan K, Dobhal H, Valliappa C, Shrestha S, and Kedam G 2019 A Machine Learning Model for Air Quality Prediction for Smart Cities *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)* Chennai India pp 452-457 doi: 10.1109/WiSPNET45539.2019.9032734
[12] Masih A 2019 Machine learning algorithms in air quality modeling Global J Environ Sci Manage 5(4) p 515-534 doi: 10.22034/gjesm.2019.04.10
[13] Xayasouk T and Lee H 2019 Air pollution prediction system using deep learning *WIT Transactions on Ecology and the Environment* vol 230 p 71-79 doi: 10.2495/AIR180071
[14] He G, Pan Y, and Tanaka T 2020 The short-term impacts of COVID-19 lockdown on urban air pollution in China *Nat Sustain* 3 p 1005-1011 doi: 10.1038/s41893-020-0581-y
[15] Navares R and Aznarte J L 2019 Predicting air quality with deep learning LSTM: Towards comprehensive models *Ecological Informatics, 101019.* doi: 10.1016/j.ecoinf.2019.101019
[16] Wibowo F W and Wihayati 2021 Prediction modelling of COVID-19 on provinces in Indonesia using long short-term memory machine learning *J. Phys.: Conf. Ser.*
[17] Jakarta Open Data 2020 Indeks standar pencemaran udara ispu tahun 2020 https://data.jakarta.go.id/dataset/indeks-standar-pencemaran-udara-ispu-tahun-2020 (accessed on 1 November 2020)
[18] The World Air Quality Project 2020 Real-time air quality monitoring in Indonesia https://aqicn.org/faq/2015-07-18/real-time-air-quality-monitoring-in-indonesia/ (accessed on 1 November 2020)
[19] AirNow 2020 Air quality index (AQI) basics https://www.airnow.gov/aqi/aqi-basics/ (accessed on 1 November 2020)
[20] Kingma D P and Ba J L 2015 ADAM: A Method for Stochastic Optimization *International Conference on Learning Representations (ICLR)* p 13
[21] Gubernur DKI Jakarta 2020 *Pelaksanaan pembatasan sosial berskala besar dalam penanganan corona virus disease 2019 (COVID-19) di provinsi daerah khusus ibukota Jakarta* Peraturan gubernur DKI Jakarta Nomor 33 Tahun 2020 (Jakarta)

Acknowledgments
The authors would like to thank Hemispheres for allowing us to publish our research paper.