

Classification and Prediction of Air Quality in Yogyakarta City Using Long Short-Term Memory Algorithm

Novita Bayu Permatasari¹, Hidayat Nur Isnianto^{1*}

¹Department of Electrical Engineering and Informatics, Universitas Gadjah Mada; novitabayu@mail.ugm.ac.id

*Correspondence: hnisnianto@ugm.ac.id

Intisari – Kualitas udara di Kota Yogyakarta mengalami penurunan seiring dengan meningkatnya pencemaran udara yang disebabkan oleh aktivitas manusia, seperti transportasi dan industri, sehingga berdampak serius terhadap kesehatan dan lingkungan. Beberapa parameter yang digunakan adalah PM10, SO2, CO, O3, dan NO2 yang didasarkan pada data ISPU (Indeks Standar Pencemar Udara) dari Dinas Lingkungan Hidup Kota Yogyakarta selama tahun 2020 hingga 2023. Dalam penelitian ini, dilakukan klasifikasi dan prediksi menggunakan algoritma LSTM (Long Short-Term Memory) untuk memprediksi standar kualitas udara berdasarkan lima parameter pencemar tersebut. Hasil penelitian pada grafik kinerja menunjukkan bahwa model cenderung mengalami overfitting pada beberapa fase pelatihan, meskipun hasil simulasi pada data acak menunjukkan konsistensi prediksi dengan kategori “Baik” untuk kualitas udara. Dengan demikian, dapat disimpulkan bahwa penelitian ini menggunakan model LSTM (Long Short-Term Memory) mampu mengklasifikasikan dan memprediksi kualitas udara secara efektif dengan hasil yang hampir sesuai dengan data aktual, serta penelitian ini akan memberikan hasil yang lebih optimal dengan beberapa perbaikan, seperti penambahan parameter PM2.5 dan peningkatan akurasi model.

Kata kunci – Kualitas udara, ISPU (Indeks Standar Pencemar Udara), Long Short-Term Memory (LSTM), PM10, SO2, CO, O3, dan NO2

Abstract – Air quality in Yogyakarta City is declining along with the increase in air pollution caused by human activities, such as transportation and industry, thus having a serious impact on health and the environment. Some of the parameters are PM10, SO2, Com O3, and NO2 which are based on ISPU (Air Pollution Standard Index) data from the Yogyakarta City Environment Agency from 2020 to 2023. In this study, classification and prediction were carried out using the LSTM (Long Short-Term Memory) algorithm to predict air quality standards with five pollution parameters of PM10, SO2, CO, O3, and NO2. The results of this study on the performance graph showed that the model was likely to be overfitted in some phases of training, although the simulation results on the random data showed consistency in the prediction with the "Good" category for air quality. So it can be concluded that this study using the LSTM (Long Short-Term Memory) model is able to classify and predict air quality effectively with results that are almost in accordance with the actual data and this study will run better with several improvements such as the use of additional parameters such as PM2.5 and increased model accuracy for optimal results.

Keywords – Air quality, ISPU (Air Pollution Standard Index), Long Short-Term Memory (LSTM), PM10, SO2, CO, O3, and NO2

I. INTRODUCTION

Human population growth is increasing every year, as is the case in Indonesia within 10 years with data obtained in 2020 there was a population increase of 1.25% which was listed on the Central Statistics Agency website which was released on January 21, 2021 [1]. With population growth, the level of air quality is also getting lower due to the increase in industrial activities, transportation, and excessive energy consumption, thus causing the emission of air pollutants such as particulate matter, greenhouse gases, sulfur dioxide (SO₂), nitrogen dioxide (NO₂) and so on.

Air is a mixture of gases that make up the earth's atmosphere that can be breathed by humans. It is mostly composed of nitrogen (78%) and oxygen (21%) and some small gases such as argon, carbon dioxide, neon, helium, methane, and others [2]. According to the WHO (World Health Organization), air quality reflects the level of atmospheric pollution by pollutants that can harm human health and the environment. Air quality is also an active research subject, and researchers from various disciplines have investigated various aspects related to air quality [3].

The impact of this air pollution on human health will be serious, such as respiratory diseases such as asthma and bronchitis, as well as cardiovascular problems because small particles such as PM 2.5 can enter the lungs and even into the bloodstream, causing harmful effects to cancer. The health impacts of air pollution can vary depending on the type of pollutant, exposure level, and individual conditions [4]. With the increasing sophistication of technology, to research about air quality can use sophistication *Artificial Intelligence* (AI). *Artificial Intelligence* or often referred to as AI, it is the ability of a computer or machine to perform tasks that usually require human intelligence.

Air quality is one of the important indicators in maintaining the balance of the environment and human health. Exposure to polluted air can have significant negative impacts. According to data from the World Health Organization (WHO), as many as 99% of the world's population lives in areas with air pollution levels that exceed safe thresholds. As a result, about seven million premature deaths occur each year due to air pollution. These harmful pollutants can come from human activities or natural factors, and spread into the atmosphere to pollute the air we breathe every day [5], [6].

In general, air quality is categorized into five levels, namely:

1. Good category with a value range of 0 – 50.
2. Medium category with a value range of 51 – 100.
3. Unhealthy Category with a score range of 101 - 199.
4. Very Unhealthy category with a score range of 200 - 299.
5. Dangerous category value range above 300 [7].

Some of the main pollutants that are of concern in air quality assessment include particulate matter (PM), carbon monoxide (CO), nitrogen dioxide (NO₂), ozone (O₃), and sulfur dioxide (SO₂) [8]. Particulate matter consists of solid particles and very small liquid droplets, including PM10 and PM2.5 which are very dangerous because they can enter the human respiratory system. These sources of particulate matter come from vehicle fumes, construction activities, and industry, among others [9].

Carbon monoxide (CO) is a colorless, odorless gas that is formed as a result of imperfect combustion of fuel. This gas is highly toxic and can be deadly at high concentrations [10]. Meanwhile, nitrogen dioxide (NO₂) is a reddish-brown gas produced from the burning of fossil fuels and industrial activities and can cause a variety of health problems as well as environmental damage [11]. In addition, ozone (O₃) in the troposphere layer acts as a harmful pollutant, although in the stratosphere it functions to protect the earth from ultraviolet radiation [12]. Sulfur dioxide (SO₂) is produced from the burning of coal and petroleum, and it causes irritation and contributes to acid rain [13]. Outdoor air pollution, both in urban and rural areas, produces fine particles that can trigger a variety of serious diseases such as stroke, heart disease, lung cancer, as well as acute and chronic respiratory distress.

As technology develops, artificial intelligence (AI) has become a very potential tool in monitoring and predicting air quality. AI falls into two main categories, namely Narrow AI designed for specific tasks, and Strong AI that includes Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI) [14]. In the field of sequential data processing such as air quality data, the Long Short-Term Memory (LSTM) algorithm is used because it is able to store and process information over long and short periods of time [15], [16]. LSTM is a form of artificial neural network (Recurrent Neural Network (RNN) that is effective for time-series data [17], [18]. To measure the performance of AI-based predictive models, evaluation methods such as the Confusion Matrix are used which show the accuracy of classification through the parameters True Positive, True Negative, False Positive, and False Negative. In addition, to understand the priority of improving the quality of services or systems, the Importance-Performance Matrix (IPM) is also used which evaluates attributes based on the level of importance and user perception performance [19], [20].

In this study, we will use the sophistication of artificial human intelligence, namely Artificial Intelligence (AI) by

using an algorithm that is tasked with classifying with the Long Short-Term Memory (LSTM) model with five parameters in this study, namely PM10, O₃, CO, NO₂, SO₂. Long Short-Term Memory is a better method compared to conventional, the Long Short Term Memory method is very suitable to be applied to sentiment analysis and LSTM requires the use of regression techniques, so that the research can be used as a reference in predicting and classifying the air quality (good or bad) of the city of Yogyakarta.

II. METHODOLOGY

In this study, several tools and materials were used that supported the success of this research. Here are the tools and materials listed on Table 1.

Table 1. Research tools and material

Tools and Materials	Function
Laptop	Used in this study, this tool functions for programming on the Google Collabs platform.
Software and Library	process instructions with desired information results
Google Collabs	As a platform for running programs as well as Writing, editing, and running <i>Python code</i>
Internet	Running programs on <i>Google Collabs</i>
Datasheet ISPU	This ISPU data is material needed in programs in training models to be able to carry out predictions and category classifications

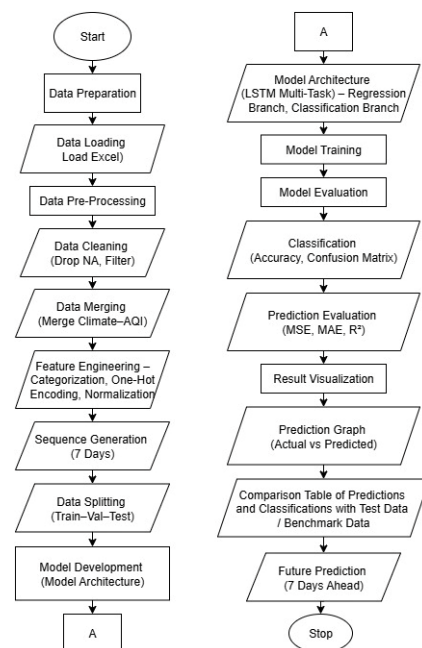


Figure 1. Research stage flowchart

The research stage begins with the determination of the research idea, followed by the preparation of the program framework according to the desired output, as well as the collection of air quality data from the Yogyakarta City Environment Agency. The flowchart in Figure 1 shows the

overall sequence of the program's processes. Figure 2 shows the flowchart of the creation of an AI program to classify air quality in Yogyakarta City using the ISPU dataset and five parameters (PM10, O3, CO, NO2, SO2). After the classification was carried out, ISPU data for the last three years was used to predict air quality one year ahead.

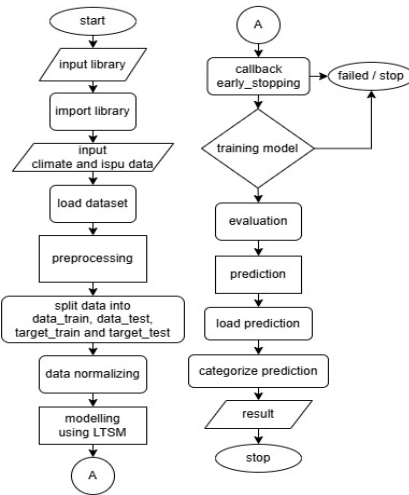


Figure 2. Process flowchart air quality classification and prediction

The system starts by importing *libraries* (pandas, numpy, random, matplotlib.pyplot, and tensorflow) into GoogleCollab. Furthermore, climate and ISPU data for 2020–2023 are imported and processed through data *cleaning* to eliminate incomplete data. As shown in Figure 3.

Bulan	: Januari 2021 s/d						
Kota	: Yogyakarta						
Stasiun	: Yogyakarta						
Waktu	Kec.Angin m/s	Arah Angin °	Kelambaban %	Suhu °C	Tek.Udara mBar	Sol.Rad watt/m ²	Curah Hujan mm
05 February 2022							
06 February 2022							
07 February 2022	2.00	12.00	87.00	26.60	996.07	141.00	0.00
08 February 2022	1.33	161.58	84.75	28.09	994.97	109.58	1.18
09 February 2022	2.08	177.75	70.08	31.36	996.23	586.33	0.00
10 February 2022	2.00	183.85	69.80	30.47	994.15	434.75	0.00
11 February 2022	2.42	183.50	68.00	31.49	994.55	637.83	0.01
12 February 2022							
13 February 2022							

Figure 3. Examples of incomplete climate data

Waktu	Kec.Angin	Arah Angin	Kelambaban	Suhu	Tek.Udara	Sol.Rad	Curah Hujan
304 2021-01-01 00:00:00	0.75	197.708333	88.895833	25.93125	29.402979	122.3125	0.15125
305 2021-01-02 00:00:00	1.208333	216.479167	80.104167	27.089583	29.415792	184.1875	0.0225
306 2021-01-03 00:00:00	1.354167	231.104167	79.958333	26.80625	29.420958	211.25	0
307 2021-01-04 00:00:00	1	167.083333	74.085714	27.974286	29.386943	270.145833	0
308 2021-01-05 00:00:00	1.083333	124.520833	81.448276	27.448276	29.385345	127.520833	0.037917

PM10	SO2	CO	O3	NO2	Critical Component
7.458333	0.00	20.791667	0.000	0.0	CO
7.458333	0.00	20.791667	0.000	0.0	CO
9.416667	1.00	24.500000	0.000	0.0	CO
10.291667	0.25	21.583333	0.000	0.0	CO
4.750000	0.00	21.125000	1.625	0.0	CO

Figure 4. Incorporation of climate and ISPU data in the program

The LSTM model was used for a prediction simulation with 10 random data to classify air quality. Figure 4 shows the results of combining climate and ISPU data in the program, thus facilitating the further analysis process.

III. RESULTS AND DISCUSSION

In the results of this model classification from the results of model training, then the model will be tested with test data so that the results of Performance Evaluation on the Training Data are obtained. The data shown in Table 2 is produced.

The imbalance in the training data results in most instances being classified as *good*. To address this, I use the F1-score as the primary evaluation metric. The F1-score balances precision and recall, making it more representative than accuracy when assessing the model's ability to perform across all categories, including those with fewer samples.

The dataset consists of three categories: *good*, *medium*, and *danger*. However, the *danger* class contains only three samples, which are insufficient for reliable validation and testing unless resampling techniques are applied.

Table 2. Performance evaluation value of training data, validation and test for classification data

Classification	Validation Data			
	P	R	F1-S	S
Good	0.95	0.95	0.95	44
Medium	-	-	-	2
Unhealthy	-	-	-	-
Very Unhealthy	-	-	-	-
Danger	-	-	-	-

Classification	Data Test			
	P	R	F1-S	S
Good	1	0.94	0.97	16
Medium	0	0	0	0
Unhealthy	-	-	-	-
Very Unhealthy	-	-	-	-
Danger	-	-	-	-

Classification	Data Train			
	P	R	F1-S	S
Good	1	1	1	542
Medium	1	1	1	19
Unhealthy	-	-	-	-
Very Unhealthy	-	-	-	-
Danger	1	1	1	3

Information:

P : Accuracy

R : Recall

F1-S : F1 Score

S : Support

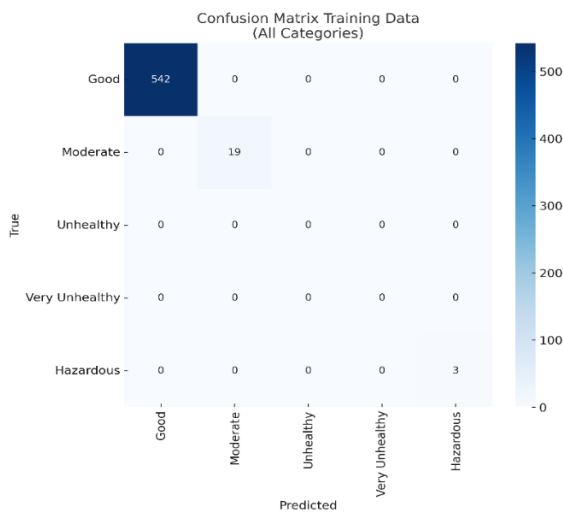


Figure 5. Confusion matrix classification data train

From the data produced, the confusion matrix will be seen from each data. In Figure 5 is a picture of the *Confusion Matrix* classification of *Data Training*, the model works very well, with 542 correct prediction data for the "Good" category and 19 data for "Medium", and 3 for "Dangerous". There are not even any misclassifications in the training data.

Figure 6 is the Confusion Matrix classification data. The Validation Model is starting to show prediction errors, although it is still quite good. For 42 samples "Good" was correctly classified, but 2 samples were incorrectly classified as "Medium" and 2 "Moderate" samples were incorrectly classified as "Good".

Figure 7 shows the *Confusion Matrix* of the test data classification results. The model correctly classified 15 samples as "Good" and 1 sample as "Moderate", with no other category predictions. Although the accuracy is high in training data, there are errors in validation and testing, which indicate potential *overfitting*. The model can still be improved by adjusting parameters or adding data.

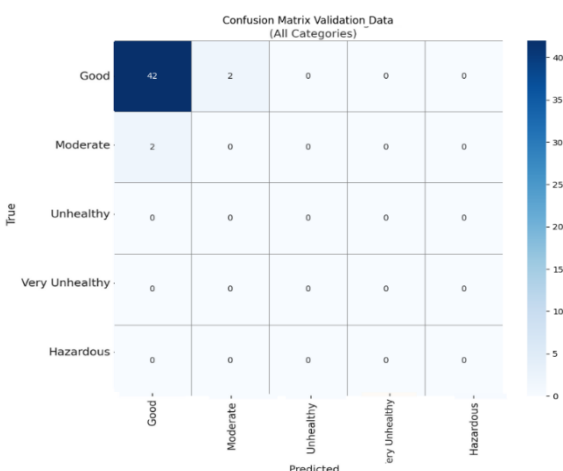


Figure 6. Confusion matrix validation data classification

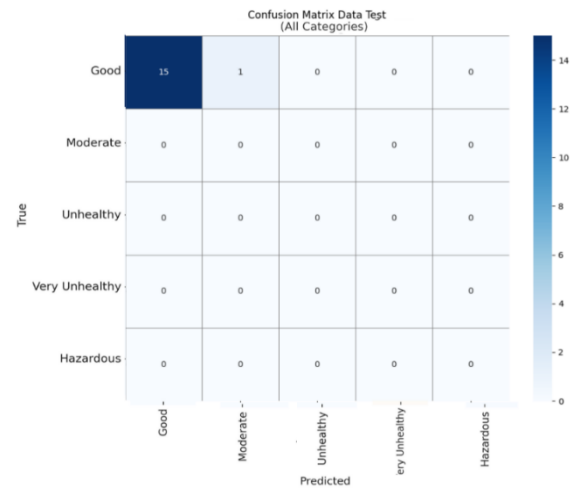


Figure 7. Confusion matrix classification data test

Table 3. Performance evaluation value of data training, validation and testing for prediction data

Pollutant	Data Train			
	MSE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	R ² (%)
PM10	6.53	2.56	1.90	92.11
SO2	7.14	2.67	1.36	94.16
CO	3.78	1.94	1.45	93.92
O3	5.32	2.31	1.48	97.72
NO2	4.06	2.02	0.98	92.51
Pollutant	Validation Data			
	MSE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	R ² (%)
PM10	29.04	5.39	3.93	70.95
SO2	100.42	10.02	5.20	35.32
CO	22.91	4.79	3.39	67.73
O3	40.76	6.38	3.17	80.88
NO2	14.76	3.84	1.75	71.69
Pollutant	Test Data			
	MSE ($\mu\text{g}/\text{m}^3$)	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	R ² (%)
PM10	29.63	5.44	4.20	-1.59
SO2	23.51	4.85	3.17	72.81
CO	16.72	4.09	3.03	63.66
O3	113.95	10.67	5.90	33.66
NO2	2.83	1.68	1.28	52.55

Information:

MSE : Mean Squares Error

RMSE : Root Mean Square Error

MAE : Mean Absolute Error

R² : R-Squared Score

Table 3 shows the evaluation of air quality prediction models for five pollutants. In the training data, the model showed good performance with low MSE, RMSE, and MAE and high R². However, performance decreased on validation and test data, indicating possible *overfitting*. Some pollutants such as

PM10 have negative R^2 on the test data, indicating poor prediction, while SO₂ and CO still show fairly good accuracy ($R^2 > 60\%$). These results are also visualized in a comparison chart of actual and predicted values.

Figure 8 shows the evaluation of the PM10 prediction model. Graph shows the model following the pattern of the actual value despite the deviation with an R^2 value of 0.92 and an MSE of 6.53. Figure 9 shows the evaluation of the SO₂ prediction model. The graph compares actual and predicted values over a time span, with the model quite well following the pattern and showing high accuracy. An R^2 value of 0.94 indicates a very strong correlation, and an MSE of 7.14.

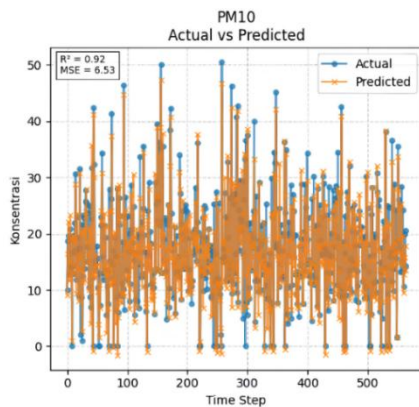


Figure 8. Graph comparison of actual values with predictions in PM10 pollutant training data

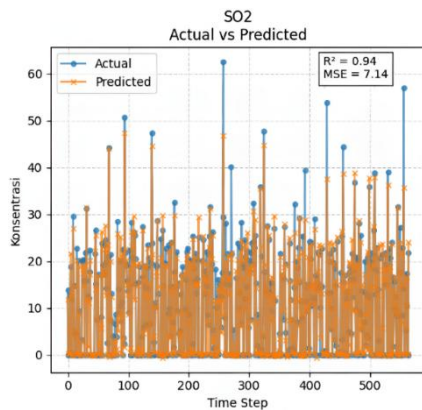


Figure 9. Graph comparison of actual values with predictions in SO₂ pollutant training data

Figure 10 shows the evaluation of the CO prediction model. The graph shows the prediction approaching the actual value despite the fluctuations, with the model able to follow the data pattern with $R^2 = 0.94$, indicating a high match, and an MSE of 3.78. Figure 11 shows the evaluation of the O₃ prediction model. The graph above shows a prediction pattern similar to the actual data although there are slight fluctuations and shows high accuracy with $R^2 = 0.98$ and MSE of 5.32.

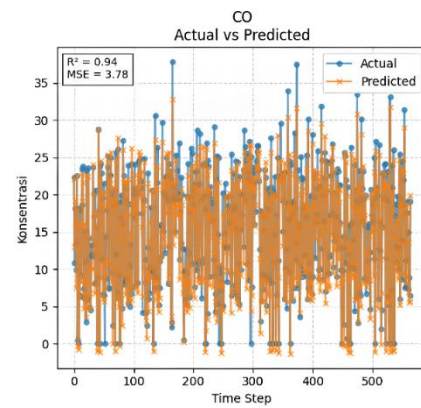


Figure 10. Graph comparison of actual values with predictions in CO pollutant training data

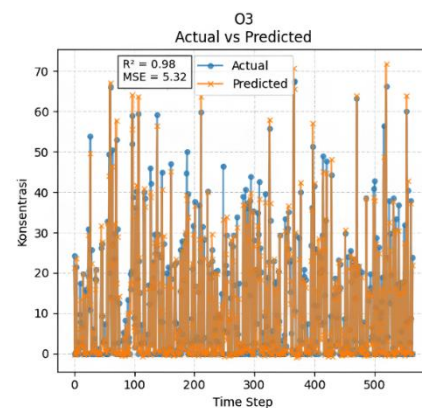


Figure 11. Graph comparison of actual values with predictions in O₃ Pollutant training data

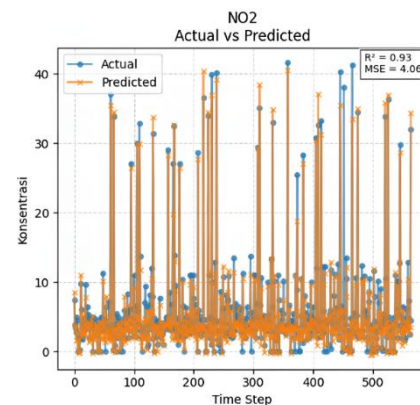


Figure 12. Graph comparison of actual values with predictions in NO₂ pollutant training data

Figure 12 shows the evaluation of the NO₂ prediction model. The graph above shows a fluctuating pattern with sharp spikes, but the prediction follows the actual data trend well with $R^2 = 0.93$ and MSE 4.06. Figure 13 shows the evaluation of the prediction model against the concentration data. The figure shows the prediction trend following the actual pattern despite the deviation with $R^2 = 0.71$ and MSE 29.04.

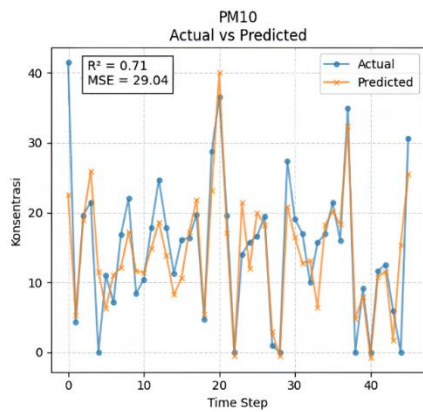


Figure 12. Graph comparison of actual values with predictions in PM10 pollutant validation data

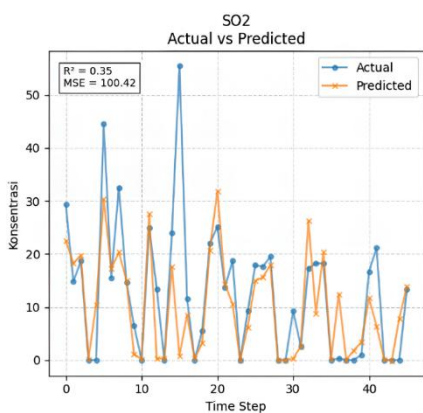


Figure 13. Graph comparison of actual values with predictions in SO2 pollutant validation data

Figure 14 shows the model's weak prediction performance. The graph shows large deviations and failure of the model to capture the peak of the actual value with $R^2 = 0.35$ and $MSE = 100.42$. Figure 15 shows the model's performance quite well, by showing the prediction following the actual pattern with a small deviation, indicating good responsiveness to the data trend with $R^2 = 0.68$ and $MSE = 22.91$.

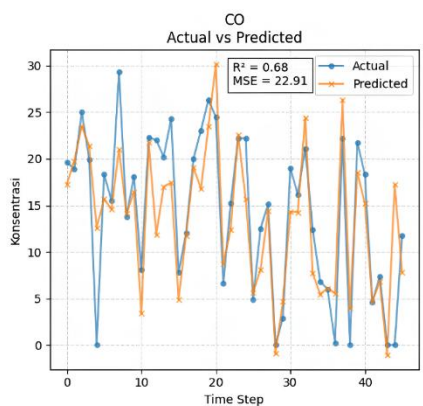


Figure 14. Graph comparison of actual values with predictions in CO pollutant validation data

Figure 16 shows that the model does quite well following actual data trends, including extreme spikes, although there are still deviations at some points. shows most predictions are close to the actual value with $R^2 = 0.81$ and $MSE = 40.76$. Figure 17 shows the model quite well following the actual data trends, although there are deviations when the spikes are sharp. With $R^2 = 0.72$ and $MSE = 14.76$.

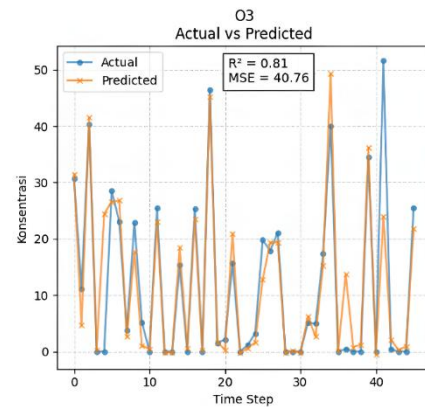


Figure 15. Graph comparison of actual values with predictions in O3 pollutant validation data

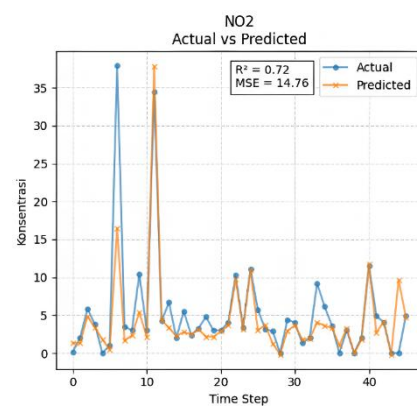


Figure 16. Graph comparison of actual values with predictions in NO2 pollutant validation data

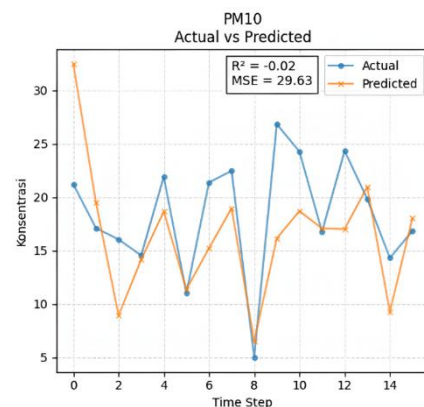


Figure 17. Graph comparison of actual values with predictions in PM10 pollutant test data

Figure 18 shows the prediction almost following the actual data trends, but there is a discrepancy. The values of $R^2 = -0.02$ and $MSE = 29.63$ indicate that the model's performance is very weak with considerable prediction errors. Figure 19 shows the prediction model can follow the general pattern of actual data, although there are significant differences at some points. With $R^2 = 0.73$ and $MSE = 23.51$.

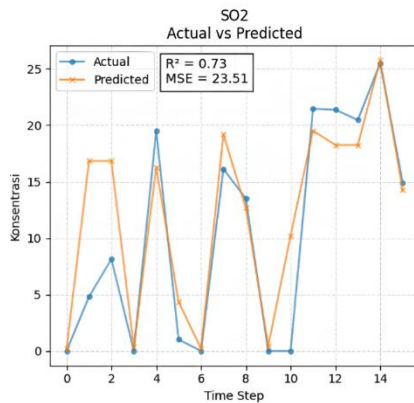


Figure 18. Graph comparison of actual values with in SO2 pollutant test data

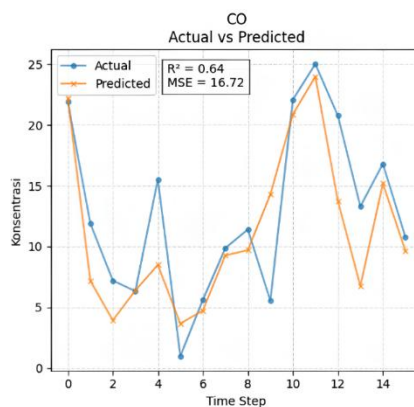


Figure 19. Graph comparison of actual values with predictions in CO pollutant test data

The graph in Figure 20 of the prediction lines tends to follow the general movement of the actual line, indicating that the prediction model has the ability to partially capture patterns and trends in the actual data. However, there are some differences between actual and predicted values. As well as producing a determination coefficient value (R^2) = 0.64 and a Mean Squared Error (MSE) value of 16.72, the MSE value provides a quantitative measure of the magnitude of the average prediction error.

The graph in Figure 21 of the prediction lines tends to follow the general movement of the actual line, indicating that the prediction model has the ability to capture some of the patterns and trends in the actual data. However, there are some differences between actual and predicted values. As well as producing a determination coefficient value (R^2)

= 0.34 and a Mean Squared Error (MSE) value of 113.95, this fairly high MSE value confirms that the prediction error of this model is relatively large.

Figure 22 shows the model's predictions generally follow the trends of the actual data even if there are discrepancies during value spikes. With $R^2 = 0.53$ and $MSE = 2.83$, the model shows a fairly low prediction error despite its moderate accuracy. The next result is the prediction result that uses the LSTM algorithm to be able to determine the air quality produced by comparing ISPU data.

Table 4 The results of the air quality prediction show that most samples are estimated to have "Good" air quality, although the highest pollutants vary (PM10, O3, SO2, CO). For example, in one sample, PM10 was predicted to be the highest and the air quality remained "Good". Similar patterns were seen in other samples with variations in predictive values.

Table 5 Data shows air pollution levels (PM10, SO2, CO, O3, NO2) for some samples. Each row is identified with its highest pollutants and air quality categories. Although the highest pollutants varied (CO, PM10, O3, SO2), most samples showed "Good" air quality.

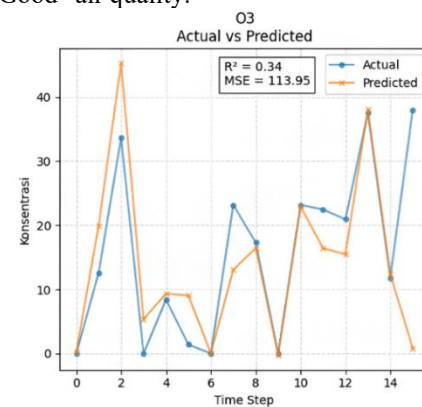


Figure 20. Graph comparison of actual values with predictions in O3 pollutant test data

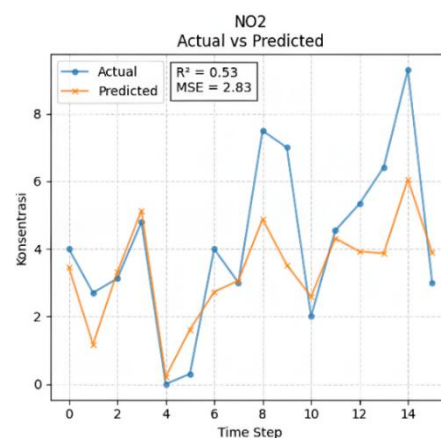


Figure 21. Graph comparison of actual values with predictions in NO2 pollutant test data

Table 4. Air quality prediction results in data test

NO	PM ₁₀	SO ₂	CO	O ₃	NO ₂	HIGHEST POLLUTANTS	CATEGORY
1	32.47	0.21	22.21	0.21	3.45	PM10	Good
2	19.50	16.83	7.20	19.96	1.18	O ₃	Good
3	8.92	16.82	3.94	45.27	3.31	O ₃	Good
4	14.15	0.32	6.38	5.38	5.13	PM10	Good
5	18.66	16.27	8.50	9.32	0.21	PM10	Good
6	11.35	4.35	3.67	9.05	1.61	PM10	Good
7	15.21	0.21	4.75	0.06	2.72	PM10	Good
8	18.95	19.19	9.27	13.05	3.06	SO ₂	Good
9	6.51	12.69	9.71	16.45	4.88	O ₃	Good
10	16.14	0.42	14.28	-0.24	3.53	PM10	Good
11	18.70	10.23	20.86	22.94	2.59	O ₃	Good
12	17.07	19.51	23.97	16.43	4.32	CO	Good
13	17.01	18.25	13.75	15.47	3.92	SO ₂	Good
14	20.96	18.24	6.76	38.07	3.87	O ₃	Good
15	9.25	25.77	15.18	12.37	6.06	SO ₂	Good
16	18.03	14.28	9.66	0.85	3.89	PM10	Good

Table 5. Air quality prediction results on actual data

NO	PM ₁₀	SO ₂	CO	O ₃	NO ₂	HIGHEST POLLUTANTS	CATEGORY
1	21.20	0.00	21.90	0.00	4.00	CO	Good
2	17.10	4.80	11.90	12.60	2.70	PM10	Good
3	16.04	8.12	7.21	33.67	3.12	O ₃	Good
4	14.54	0.00	6.33	0.00	4.79	PM10	Good
5	21.90	19.50	15.50	8.40	0.00	PM10	Good
6	11.00	1.00	1.00	1.40	0.30	PM10	Good
7	21.38	0.00	5.62	0.00	4.00	PM10	Good
8	22.46	16.12	9.88	23.17	3.00	O ₃	Good
9	5.00	13.50	11.40	17.30	7.50	O ₃	Good
10	26.83	0.00	5.54	0.00	7.00	PM10	Good
11	24.27	0.00	22.07	23.16	2.00	PM10	Good
12	16.75	21.46	25.00	22.46	4.54	CO	Good
13	24.33	21.38	20.75	20.92	5.33	PM10	Good
14	19.83	20.46	13.29	37.50	6.42	O ₃	Good
15	14.30	25.50	16.80	11.80	9.30	SO ₂	Good
16	16.83	14.92	10.79	37.95	3.00	O ₃	Good

Table 6 shows the accuracy of the highest pollutant prediction is only 50%, while the air quality category is quite accurate with 1 error out of 14 data. This indicates that the model is more reliable at predicting categories than dominant pollutants, so improvements are needed in this aspect.

For Figure 23 The seven-day pollutant prediction graph shows changes in the concentration of five major pollutants: PM10, SO2, CO, O3, and NO2. The horizontal axis indicates the 0th to 6th day, and the vertical axis indicates the concentration of pollutants. PM10 is predicted to rise until the 5th day and then decrease slightly. SO2 rose sharply until the 2nd day, then decreased. CO increases steadily until day 4 and is relatively stable thereafter. O3 fluctuates, while NO2 tends to be stable with little change. This graph helps monitor potential changes in air quality.

Table 6. Comparison of results from actual data with predictive data

Highest pollutants actual data	Actual data categories	Highest pollutants predicted data	Prediction data categories	Conformance of Pollutants	Category Fit
CO	Good	PM10	Good	FALSE	TRUE
PM10	Good	O ₃	Good	FALSE	TRUE
O ₃	Good	O ₃	Keep	TRUE	FALSE
PM10	Good	PM10	Good	TRUE	TRUE
PM10	Good	PM10	Good	TRUE	TRUE
PM10	Good	PM10	Good	TRUE	TRUE
O ₃	Good	SO ₂	Good	FALSE	TRUE
O ₃	Good	O ₃	Good	TRUE	TRUE
PM10	Good	PM10	Good	TRUE	TRUE
PM10	Good	O ₃	Good	FALSE	TRUE
CO	Good	CO	Good	TRUE	TRUE
PM10	Good	SO ₂	Good	FALSE	TRUE
O ₃	Good	O ₃	Good	TRUE	TRUE
SO ₂	Good	SO ₂	Good	TRUE	TRUE
O ₃	Good	PM10	Good	FALSE	TRUE

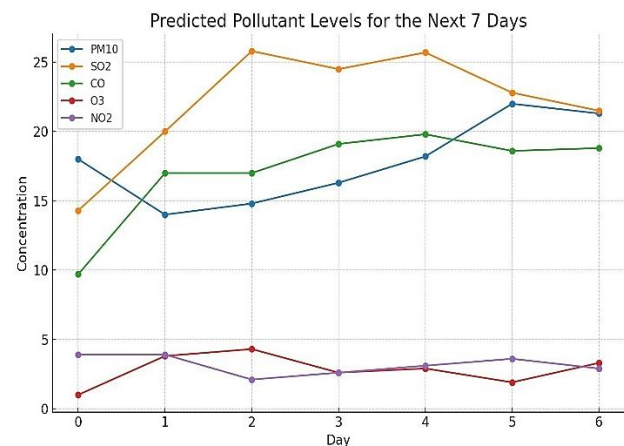


Figure 22. Sample pollutant level prediction chart for the next 7 days

IV. CONCLUSION

Based on this study, there are several conclusions, including the following:

1. The results of this study succeeded in producing a category classification on air quality with the category "Good" in the sample made, so that the air quality is still within safe limits in accordance with air quality standards. There are no concentration of pollutant parameters that are too high values that can indicate that the air quality is sufficiently consistent for all observations. The accuracy in this study was 93.75%.
2. The results of the predictions made are almost in accordance with the actual value obtained and there are some systematic errors that are not so significant, so that they can be used as a reference to find out the air quality in the future.

In this study, it is recommended to add more data to make the category distribution more representative, if imbalance still occurs, apply data augmentation such as oversampling, ensure that the data is split evenly so all categories are represented in both the training and test sets, and apply various cross-validation methods to ensure a fair and consistent model evaluation.

REFERENCES

- [1] H. Fuadi, B. Ismiwati and B. Saripta, "Peningkatan Laju Pertumbuhan Penduduk Terhadap Pembangunan Keluarga," *Empiricism Journal*, vol. 4, no. 2, pp. 673-678, 2023.
- [2] S. K. M. Anwar Mallongi, "PENCEMARAN DARI INDUSTRI DAN ANALYSIS RISIKO EKOLOGI".
- [3] V. Rahmawati, A. L. Hayat and A. Salam, "Analisis Dampak Pencemaran Udara Terhadap Kesehatan Masyarakat Di Perkotaan," *Jurnal Sosial dan Pengabdian Masyarakat*, vol. 2, no. 3, 2024.
- [4] A. A. Anandari, A. F. Wajdi and G. Harsono, "Dampak Polusi Udara terhadap Kesehatan dan Kesiapan Pertahanan Negara di Provinsi DKI Jakarta," *Journal on Education*, vol. 06, no. 2, pp. 10868-10884, 2024.
- [5] J. T. Santoso, "Kecerdasan Buatan (Artificial Intelligence)," *Penerbit Yayasan Prima Agus Tek.*, pp. 1-227, 2023.
- [6] E. S. Eriana and A. Zein, *Artificial Intelligence (AI). Eureka Media Aksara*. 2023.
- [7] C. N. Daiman, A. Y. Rahman and F. Nudiyansyah, "Klasifikasi Teks Berita Breaking News Di Manggarai Menggunakan Long Short Term Memory (LSTM)," *Jurnal MNEMONIC*, vol. 7, no. 2, pp. 170-174, 2024.
- [8] A. E. Putra and T. Rismawan, "Klasifikasi Kualitas Udara Berdasarkan Indeks Standar Pencemaran Udara (ISPU) Menggunakan Metode Fuzzy Tsukamoto," *Jurnal Komputer dan Aplikasi*, vol. 11, no. 2, pp. 190-196, 2023.
- [9] F. Haya, K. Nisa, R. F. Ladipasa, A. Suriani and A. Media, "Dampak Polusi Udara terhadap Kesehatan Manusia," *Jurnal Ilmu Sosial dan Humaniora*, vol. 3, no. 2, pp. 180-190, 2025.
- [10] K. D. Pertiwi, I. P. Lestari and A. Afandi, "Analisis Risiko Kesehatan Lingkungan Paparan Debu PM10 dan PM2.5 pada Relawan Lalu Lintas di Jalan Diponegoro Ungaran," *Pro Health Jurnal Ilmiah Kesehatan*, vol. 6, no. 2, pp. 85-91, 2024.
- [11] D. A. Putri, A. Rosyada, W. Lionita and F. Hepiman, "Gangguan Kesehatan Akibat Paparan Karbon Monoksida pada Penjual Sate di Pinggir Jalan," *Jurnal Ilmu Kesehatan Masyarakat*, vol. 13, no. 2, pp. 123-130, 2024.
- [12] P. Vongelis, N. G. Koulouris, P. Bakakos and N. Rovina, "Air Pollution and Effects of Tropospheric Ozone (O3) on Public Health," *International Journal of Environmental Research and Publik Health*, vol. 22, p. 709, 2025.
- [13] S. Amalia and I. R. Wahyuni, "Analisis Sulfur Dioksida (SO2) Udara Ambient Menggunakan Metode Pararosanilin dengan Spektrofotometer UV-Visible Kabupaten Bandung, Jawa Barat," *Gunung Djati Conference Series*, vol. 15, 2022.
- [14] R. S. Y. Zebua *et al.*, *Fenomena Artificial Intelligence (AI)*. PT. Sonpedia Publishing Indonesia, 2023.
- [15] M. F. Naufal and S. F. Kusuma, "Analisis Perbandingan Algoritma Machine Learning dan Deep Learning Untuk Klasifikasi Citra Sistem Isyarat Bahasa Indonesia (SIBI)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 873-882, 2023.
- [16] R. Gunawan and R. Wesley, "Literatur Review: Metode Deep Learning Untuk Analisis Teks," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 5, 2025.
- [17] J. Cahyani, S. Mujahidin and T. P. Fiqar, "Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional," *Jurnal Sistem dan Teknologi Informasi*, vol. 11, no. 2, 2023.
- [18] S. Y. Prasetyo and G. Z. Nabiilah, "Perbandingan Model Machine Learning Pada Klasifikasi Tumor Otak Menggunakan Fitur Discrete Cosine Transform," *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 29-34, 2023.
- [19] N. Slack, "The Importance-Performance Matrix as a Determinant of Improvement Priority," *International Journal of Operations & Production Management*, vol. 14, no. 5, pp. 59-75.
- [20] A. Yani, "Peran Artificial Intelligence sebagai Salah Satu Faktor dalam Menentukan Kualitas Mahasiswa di Era Society 5.0," *Journal of Education Research*, vol. 5, no. 2, pp. 1089-1096, 2024.