

Anomaly Detection in Time Series Data of VOC (Volatile Organic Compound) To Generate Indoor Air Quality Alerts

Hadi Dharma Nusantara*, Bayu Erfianto

School of Computing, Informatics, Telkom University, Bandung, Indonesia

Email: ^{1,*}hadidharma@student.telkomuniversity.ac.id, ²erfianto@telkomuniversity.ac.id

Correspondence Author Email: hadidharma@student.telkomuniversity.ac.id

Submitted: 02/02/2023; Accepted: 27/03/2023; Published: 31/03/2023

Abstract—Indoor air quality is an especially important factor and needs to be considered for health. Poor indoor air quality can trigger illness, reduce productivity, and disrupt the comfort of people in the space. In residential areas, hospitals, schools, nursing homes, and other specialized environments, indoor air pollution can affect groups that are more vulnerable to health problems due to their health conditions or age. This research aims to predict indoor air quality using the Long Short-Term Memory (LSTM) method and provide alerts when the prediction results exceed a predetermined limit. The accuracy level is measured using Mean Absolute Percentage Error (MAPE) by calculating the difference between the original data and the prediction results. In this study, a system was created that utilizes Internet of Things (IoT) technology that can monitor the state of indoor air quality such as temperature, TVOC, CO₂, and HCHO gas levels. The system uses the WP6003 Air Box Reader tool as an indoor air quality detector that is connected to the website created. This website can display data that is being recorded, download datasets that have been recorded, visualize predictions of temperature, TVOC, CO₂, and HCHO, and notify if any data crosses a predetermined limit. The results obtained are quite good prediction accuracy by getting a MAPE value of 0.30452, RMSE 0.023475, and the average value of the test data is 24.035 which means that if the RMSE value is close to 0, the prediction results will be more accurate. Anomalies result in values of room temperature and HCHO that are above normal limits.

Keywords: Air Quality; Iot; WP6003 Air Box Reader; Accuracy; LSTM

1. INTRODUCTION

Indoor air quality is one of the most important factors for health. Poor indoor air quality can trigger disease, reduce productivity and disturb people's comfort in the space [1]. In residential areas, hospitals, schools, nursing homes, and other specialized environments, indoor air pollution can affect groups more vulnerable to health problems due to their health conditions or age [2]. Indoor air quality can be tested from many parameters, some of which are volatile organic compounds (VOCs), carbon dioxide (CO₂), and formaldehyde (HCHO). (TVOC), carbon dioxide (CO₂), and formaldehyde (HCHO). These parameters can cause adverse health effects if they do not meet quality standards. TVOCs can cause headaches, nausea, liver damage, kidney damage, eye irritation, nose irritation, throat irritation, and so on [3]. CO₂ can cause arrhythmia, convulsions, and anesthesia [4].

In Indonesia, the quality standard of CO₂ in the room is regulated in the Regulation of the Minister of Health of the Republic of Indonesia No. 1007/MENKES/PER/V/2011. Indonesia number 1007/MENKES/PER/V/2011, the concentration of CO₂ in the room should not exceed 1000 ppm/8h [5]. The Occupational Safety and Health Administration (OSHA) has issued a regulation prohibiting employees from being in rooms with oxygen concentrations below 19.5% and above 22% [6]. Hong Kong Government (GovHK) issued a quality standard for TVOC, 0.6 mg/m³ per eight hours.

The importance of indoor air quality has prompted researchers' interest in air quality research. In 2015, a study of indoor air quality was conducted using the parameters (PM₁₀) and carbon dioxide (CO₂) [4] [16]. This study showed an increase in CO₂ concentration, reaching 2063 ppm when there were students in the classroom. In 2017, further research was conducted on indoor air quality based on three parameters PM_{2.5}, CO₂, and VOCs in seven rooms. The results found that the average CO₂ concentration from eight hours in seven out of eight rooms exceeded 1000 ppm, while three out of four rooms had average concentrations of PM_{2.5} of more than 35µgm⁻³ [7] [1].

Monitoring indoor air quality is important to ensure that indoor air remains clean and healthy for its occupants. Indoor air quality can be affected by numerous factors, such as air pollution from the outside, chemicals contained in household products, and human activities such as smoking. Indoor air quality monitoring is carried out to measure the level of air cleanliness and determine the actions that need to be taken to improve air quality. Indoor air quality monitoring also helps to identify sources of air pollution and determine the actions that need to be taken to address them. Therefore, to facilitate monitoring indoor air quality, a system connected to the device is needed. Connecting to the device can facilitate monitoring anywhere and anytime as long as Bluetooth connects it. For the potential for gas concentrations in the room that does not meet quality standards, indoor air quality monitoring is needed. To facilitate the monitoring of indoor air quality, a system connected to the device is needed. Connecting to the device can facilitate monitoring anywhere and anytime if Bluetooth connects it. By Bluetooth. Based on these things, the author makes an application to facilitate monitoring indoor air quality based on four parameters. Based on four parameters: TVOC, CO₂, HCHO, and temperature.

In the field of technology, the prediction of indoor air quality using machine learning methods is an effective solution for monitoring and predicting indoor air quality. One widely used machine learning method is Long Short-Term Memory (LSTM).

LSTM is an artificial neural network developed to solve signal-processing problems with long-term memory. This network is very useful in processing signals with complex temporal (time) relationships and require information from the past to predict the future [19]. LSTM can retain long-term information and limit nonessential input, making them especially useful for signal-processing applications [18].

In indoor air quality monitoring, LSTM is used to predict indoor air pollution levels based on data taken over a certain period of time. The LSTM model will study the relationship between factors affecting air quality, such as the number of occupants, activity, and outdoor air pollution, to predict future levels of indoor air pollution.

Thus, the LSTM method has the potential to be an effective solution in monitoring and predicting indoor air quality. Accurate predictions can help in taking the right actions to improve indoor air quality and maintain the health and comfort of the occupants.

2. RESEARCH METHODOLOGY

2.1 Research Stages

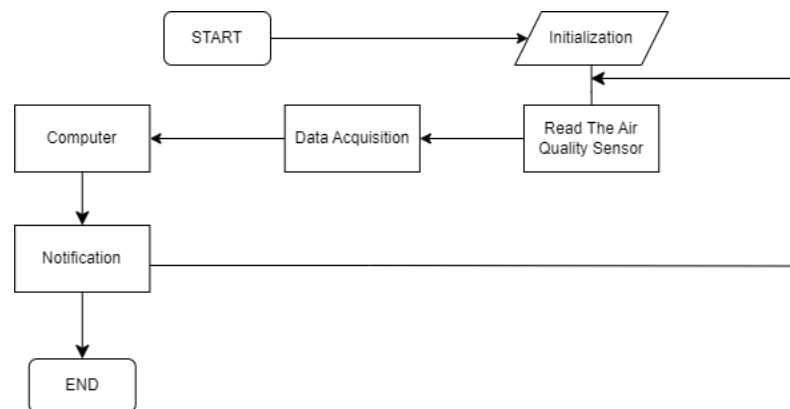


Figure 1. System Design

Figure 1 is the workings of the designed system starting from the WP6003 Air Box Reader device which functions to read and measure indoor air quality. The parameters measured are temperature, TVOC, HCHO, and CO₂. After the WP6003 Air Box Reader device is turned on, the device will measure air quality. The website that has been created will be connected to the device. The device cannot send data air quality measurement results in real-time because the device is only equipped with a Bluetooth module that cannot connect to a Bluetooth module that cannot be connected to an internet connection. The website can only record and store data that has been recorded by Air Box Reader WP6003 and can be downloaded in CSV file format. The downloaded file will be trained to predict air quality values using the Long Short-Term Memory (LSTM) algorithm model. This algorithm model can predict the possibility that will occur in the future. The LSTM model training process is conducted on the Google Collabs platform. The resulting prediction results will be uploaded back to the website so that air quality prediction data can be seen and observed by users.

2.2 Overview of The System

Figure 2 is an overview of the system containing the overall system design that will be used in the research conducted. System design used in this study:

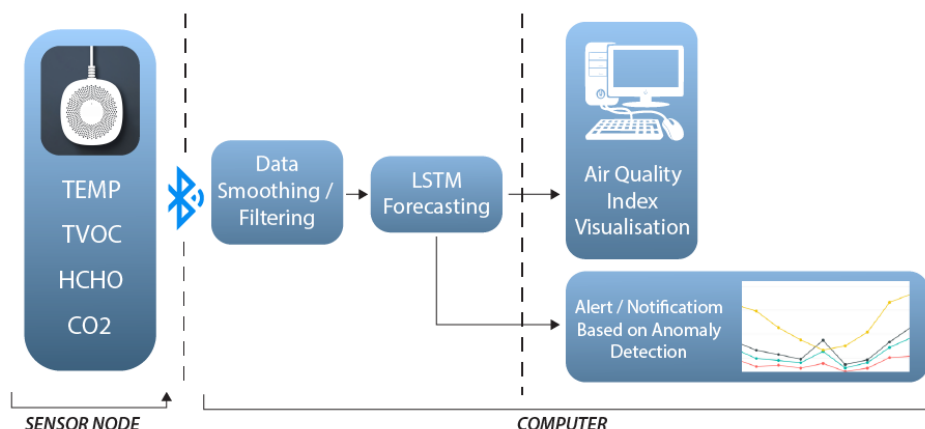


Figure 2. Overview of The System

- a. WP6003 Air Box Reader
The WP6003 air box reader is used as a tool to capture, collect, and prepare data and process it to generate the desired data.
- b. Data Smoothing/ Filtering
Filtering and smoothing the data containing noise that has been obtained from the WP6003 Air Box Reader to produce smoother data processing and smooth curves.
- c. LSTM Forecasting
Data prediction using short-term long-term memory method complements the shortcomings of recurrent neural network (RNN) because it cannot predict data based on past data of the recurrent neural network (RNN). After all, it cannot predict data based on past data stored for a long time.
- d. Air Quality Index Visualization
index that displays a visualization of the four predicted air quality parameters.
- e. Notification
Notification is output when there are anomalies of the four air quality parameters in the form of a table under the visualization index.

2.3 Tool Design

The hardware design that will be designed based on Figure 3 is connecting the WP6003 air box reader with the device.

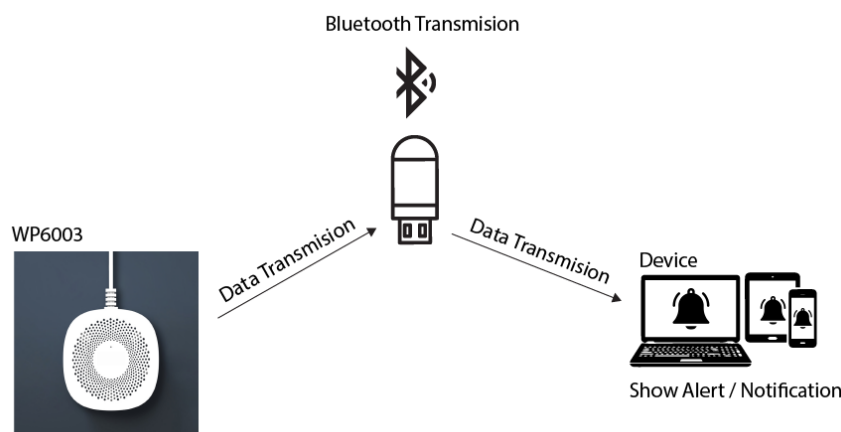


Figure 3. Tool Design

WP6003 reader with the device, the WP6003 air box reader, will read the results of the data that has been recorded, then sent via Bluetooth network and become raw data that the program will process. The following is the hardware design.

2.4 Dataset

The data collection stage in this research is taken from the recording results in the house. Data features taken from the recording results are Number, Time, Temperature, TVOC, HCHO, and CO2. Range dataset time range from November 19, 2022, to December 20, 2022, with a data retrieval frequency of every 5 seconds. data every 5 seconds and obtained 204582 rows of data.

Table 1. Dataset

No	Time	Temperature	TVOC	HCHO	CO2
0	2022-11-19 10:35:41	25.8	0.0	0.0	544
1	2022-11-19 10:35:46	25.8	0.0	0.0	544
...
204581	2022-12-20 13:47:12	26.1	0.062	0.0	537
204582	2022-12-20 13:47:17	26.1	0.064	0.0	537

The dataset used for anomaly checking uses the same data features as Table 1. The time span of the dataset is from January 23, 2023, to January 24, 2023, with a frequency of data retrieval every 5 seconds, and 13946 rows of data were obtained.

Table 2. Dataset

No	Time	Temperature	TVOC	HCHO	CO2
1	1/23/2023 05:42:37	26.5	0	0	537
2	1/23/2023 05:42:43	26.5	0	0	537

...
13945	1/24/2023 11:52:48	27.5	0.08	0.011	589
13946	1/24/2023 11:52:52	27.5	0.08	0.011	589

2.5 TVOC

Volatile Organic Compounds (VOCs) encompass a broad range of chemicals, some of which may pose harmful effects on individuals both in the short and long term. When exposed to sunlight, VOCs can interact with Nitrogen Oxides (NOx) to create ground-level ozone and smoke or haze. If the ozone concentration in the air reaches certain levels, it can harm both human health and the environment. [3].

2.6 CO2

Carbon Dioxide (CO₂), also known as carbonic acid, is a chemical composed of a carbon atom covalently bonded with two oxygen atoms. In normal temperature and pressure conditions, CO₂ is a gas. The levels of CO₂ serve as an indicator of the efficiency of a ventilation or room air system. Ideally, the CO₂ content in a room should be below 1,000 parts per million (ppm). If the CO₂ level exceeds this limit, it may mean that the ventilation system is not providing enough fresh air. [8] [20].

2.7 Temperature

Temperature is a crucial element, regarded as significant due to its dual impact: first, the temperature changes throughout the day with multiple values during the evening, night, morning, and daytime, causing alterations in underwater physical characteristics that can impact the sound signal's strength, and second, all chemical reactions, including salinity, are influenced by temperature [9].

2.8 Long-Short-Term Memory

LSTM, or Long-Short Term Memory, is a form of machine learning based on the recurrent neural network (RNN) method that can predict the current state of the machine through massive computation [2]. As an adaptive architecture neural network, its shape can be tailored to suit the specific application. A recurrent neural network is a type of network that is specifically optimized to handle sequential data [10].

3. RESULT AND DISCUSSION

This research uses a dataset that has collected as much as 204582 data, air quality obtained indoors by measuring the temperature of TVOC, HCHO, and CO₂ using the Air Box tool.

3.1 Test Results

3.1.1 Dashboard

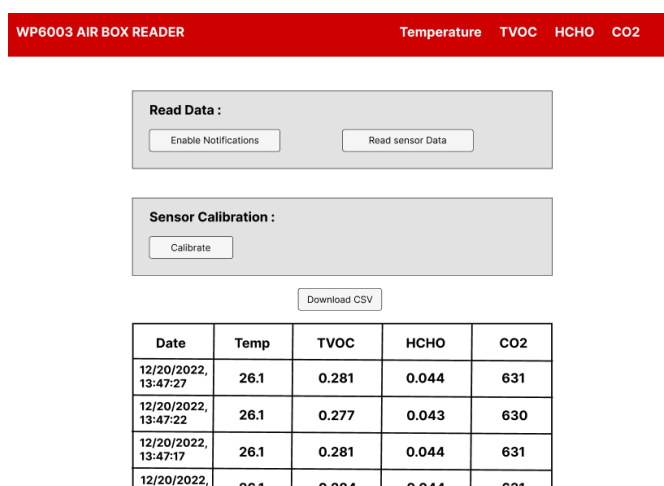


Figure 4. Dashboard

Dashboard view of the designed website. In the dashboard, there is a notifications button to activate notifications that will appear when an anomaly is detected, the read sensor data button functions to read and display data in the table from the WP6003 air box reader device, the calibrate button functions to calibrate the sensor to get the desired data and the CSV download button functions to download data that has been stored on the website in CSV data format. In the navigation bar above there are temperature, TVOC, HCHO, and CO₂ menus that function to move the page to the respective prediction pages. The test results obtained that each menu on the dashboard page can function properly and there is no missing data from the sensor readings.

3.1.2 Temperature Prediction Results

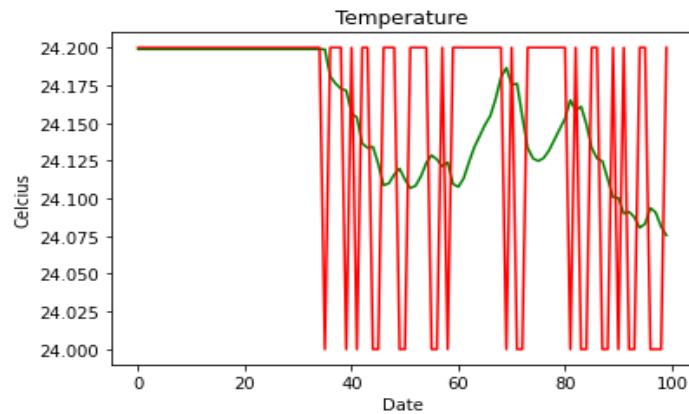


Figure 5. Temperature Prediction

Temperature prediction results obtained from the sensor reading data training process. The data displayed is the first one hundred data after the first day's data. From the picture above the red line is the original data from the device and the green line is the prediction result of the LSTM model. The results obtained using the LSTM model are good enough. In the temperature menu, there is a table that holds notifications if there are anomalies from the prediction results that exceed the predetermined limits, namely the temperature is between 18 - 30 degrees Celsius. If there is no data below 18 or above 30 then there is no data entered into the notification table.

3.1.3 TVOC Prediction Results

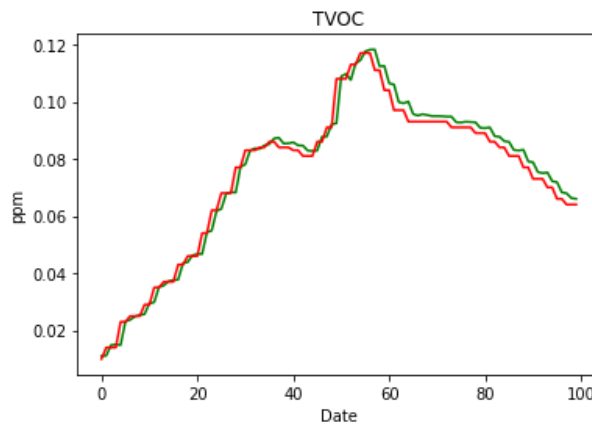


Figure 6. TVOC Prediction

TVOC prediction display obtained from training results with the same period as the temperature. The red line is the original data, and the green line is the prediction result which has a similar pattern to the red line. The same pattern as the red line. This means that the graph is incredibly good because the range of value differences does not differ much from the original data. The limit of TVOC levels is 3 ppm. If it has passed 3 ppm the data will enter the notification table because it has exceeded the reasonable limit.

3.1.4 HCHO Prediction Results

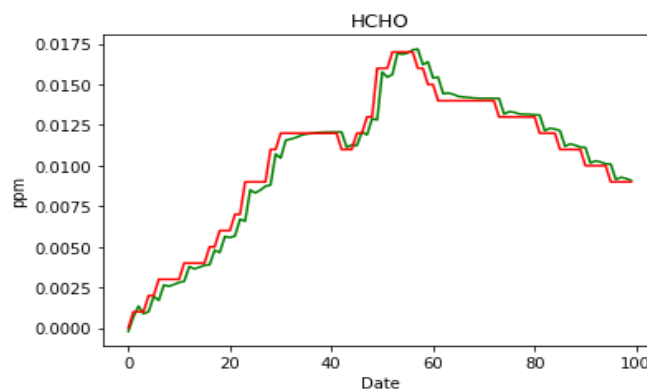


Figure 7. HCHO Prediction



HCHO prediction was obtained from training results with the same period as temperature. The red line is the original data, and the green line is the prediction result. The patterns of the original data and the prediction data are very similar and the distance between the two data is not too far, indicating that the accuracy of the prediction results is very good. The maximum level of HCHO is 0.1 ppm. If it has exceeded that, the data will be displayed in the notification table.

3.1.5 CO2 Prediction Results

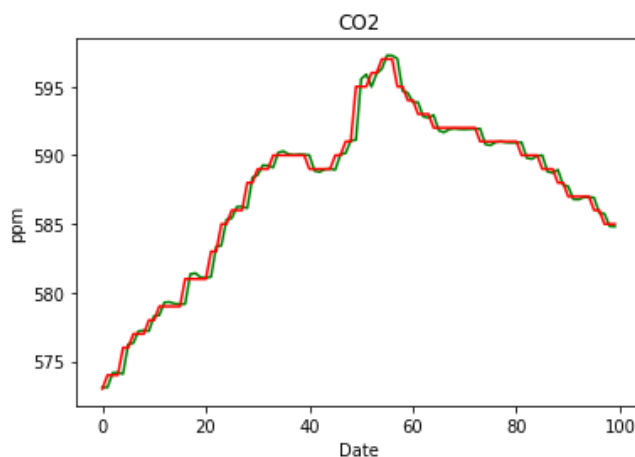


Figure 8. CO2 Prediction

CO2 readings obtained from training results with the same period as temperature. the same period as temperature. From the picture above, the two patterns of the original data values with predicted data have a pattern that is the same and the vulnerability of the difference in values obtained is not far enough so that the system prediction results are good enough. far enough so that you can say the prediction results of the system are good enough. The maximum CO2 level is below 1000 ppm. If the CO2 level has touched 1000 or more, the data will appear in the notification table.

3.1.6 Accuracy of Indoor Temperature Prediction

Long Short-Term Memory (LSTM) machine, In the test using Adam's optimization with a learning rate of 0.001 by using an epoch value of 20. The results obtained from the Adam optimization algorithm are the MAPE value at a temperature of 0.30452, the RMSE value of 0.08019, and the average value of the test data is 24.035.

a. Anomaly

Table 3. Anomaly Temperature

	Time	Temp
4865	1/23/2023 13:17:16	30.1
4869	1/23/2023 13:17:36	30.1
4871	1/23/2023 13:17:46	30.1
4872	1/23/2023 13:17:51	30.1
4873	1/23/2023 13:17:56	30.1
...
5956	1/23/2023 15:02:41	30.1
5957	1/23/2023 15:02:46	30.1
5958	1/23/2023 15:02:51	30.1
5959	1/23/2023 15:02:56	30.1
5960	1/23/2023 15:03:01	30.1

Table 4. Anomaly HCHO

	Time	HCHO
5754	1/23/2023 14:35:50	0.131
5755	1/23/2023 14:37:36	0.138
5756	1/23/2023 14:37:50	0.133
5757	1/23/2023 14:38:50	0.124

The results obtained from forecasting data, namely, at room temperature, some data is above the normal limit, precisely on January 23, 2023, at 15:03:01 getting a result of 30.1 degrees Celsius where the room temperature should be between 18 - 30 degrees Celsius. HCHO on January 23, 2023, at 14:38:50 got a result of 0.124 ppm where the

HCHO content in the room should be below 0.1 ppm. TVOC and CO₂ did not get results that exceeded their normal limits.

4. CONCLUSION

Based on the results of the research that has been done, conclusions can be drawn from this research is a web display that aims to make it easier to find out indoor air quality anomalies. The problem found from this research is that the website cannot predict in real-time, it is only limited to predicting data records that have been recorded some time back. Air quality anomalies can function properly, but we did not manage to get air quality data that could be categorized as anomalies due to the factor of the data recording place. This shows that the data recording place is feasible because the air quality is at a normal level. The second dataset is our recording results in the smoking room. With that, we can see whether the anomaly detector function is working or not. the result is that the room temperature and HCHO values are above normal limits with a room temperature value of 30.1 degrees Celsius which should be a normal temperature below 30 degrees Celsius and HCHO 0.124 ppm which should have HCHO levels below 0.1 ppm. Checking the accuracy of room temperature gives satisfactory results because the MAPE value is 0.30452, RMSE is 0.023475 and the average test data is 24.035 which means that the prediction results are more accurate if the RMSE value is close to zero. RMSE is one of the criteria for determining the forecasting model in addition to MAPE, MAD, and MSE. A low RMSE value means that the variability of the value produced by the prediction model is close to the variability of the observed value. Further research is needed to create a more effective model to improve the prediction results of the system built.

REFERENCES

- [1] A. Datta, R. Suresh, A. Gupta, D. Singh, and P. Kulshrestha, "Indoor air quality of non-residential urban buildings in Delhi, India," *International Journal of Sustainable Built Environment*, vol. 6, no. 2, pp. 412–420, Dec. 2017, doi: 10.1016/j.ijsbe.2017.07.005.
- [2] "World Health Organization Regional Office for Europe SELECTED POLLUTANTS." [Online]. Available: www.euro.who.int
- [3] Dewi W, Raharjo M, Wahyuningsih N "Literatur Review: Hubungan Antara Kualitas Udara Ruang Dengan Gangguan Kesehatan Pada Pekerja" 2021.
- [4] Majd E, McCormack M, Davis M, Curriero F, Berman J, Connolly F, ... Koehler K "Indoor air quality in inner-city schools and its associations with building characteristics and environmental factors" 2018.
- [5] "PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA."
- [6] D. Michaels and G. R. Wagner, "Occupational Safety and Health Administration (OSHA) and Worker Safety During the COVID-19 Pandemic," *JAMA*, vol. 324, no. 14, p. 1389, Oct. 2020, doi: 10.1001/jama.2020.16343.
- [7] R. N. Lesmana and Y. Rahayu, "Membangun Sistem Pemantau Kualitas Udara Dalam Ruang Dengan Mengaplikasikan Sensor CO, O₃, PM₁₀ Berbasis LabVIEW," 2018.
- [8] J. Inovasi Pendidikan Fisika dan Riset Ilmiah et al., "ANALISIS KANDUNGAN CO₂ DENGAN SENSOR DAN BERBASIS LOGGER PRO DI DAERAH YOGYAKARTA," *STKIP Nurul Huda JIPRI*, vol. 1, no. 1, pp. 5–8, 2017, [Online]. Available: <http://www.vernier.com>.
- [9] A.-B. KaishunnWu, "SPRINGER BRIEFS IN COMPUTER SCIENCE Machine Learning Modeling for IoT Networks Internet of Underwater Things." [Online]. Available: <http://www.springer.com/series/10028>.
- [10] L. Wiranda and M. Sadikin, "PENERAPAN LONG SHORT TERM MEMORY PADA DATA TIME SERIES UNTUK MEMPREDIKSI PENJUALAN PRODUK PT. METISKA FARMA." 2019.
- [11] Suharti N, Br. Surbakti "PEMANFAATAN TANAMAN (*Dieffenbachia* spp) SEBAGAI PENURUNAN ANGKA MIKROORGANISME DI UDARA PADA RUANG TUNGGU TERMINAL BUS ALS AMPLAS MEDAN" 2020.
- [12] Srivastava N, Negi S "Volatile Organic Compounds: The Concealed Depreciator of Indoor Air Quality" 2022.
- [13] WHO "Methods for sampling and analysis of chemical pollutants in indoor air" 2020.
- [14] Astuti, I. A & Firdaus, T. "ANALISIS KANDUNGAN CO₂ DENGAN SENSOR BERBASIS LOGGER PRO DI DAERAH Yogyakarta" 2017.
- [15] L. Wiranda and M. Sadikin, "PENERAPAN LONG SHORT TERM MEMORY PADA DATA TIME SERIES UNTUK MEMPREDIKSI PENJUALAN PRODUK PT. METISKA FARMA." 2019.
- [16] Sutarna N, Fatah M, Radila Z, Wicaksana H "Sistem Pengukuran Kualitas Udara (CO, O₃, PM₁₀) Jarak Jauh Berbasis Arduino Dan Android Studio" 2022.
- [17] KaishunnWu A "SPRINGER BRIEFS IN COMPUTER SCIENCE Machine Learning Modeling for IoT Networks Internet of Underwater Things" 2021.
- [18] Yu Y, Si X, Hu C, Zhang J "A Review of Recurrent Neural Networks: LSTM cells and Network Architectures" 2019.
- [19] Elmaz F, Eyckerman R, Casteels W, Latré S, Hellinckx P "CNN-LSTM architecture for predictive indoor temperature modeling" 2021.
- [20] Andrizal, Yani P I, Antonisfia Y "MONITORING DAN KONTROL KADAR CO₂ DALAM RUANGAN BERBASIS SISTEM PENCIUMAN ELEKTRONIK" 2020.