Predictive Analytics for Carbon Dioxide Levels: A Linear Regression Approach

Jasmine Sarah Maqnolia[1*], Dodik Arwin Dermawan[2], Hafizhuddin Zul Fahmi[3], I Gde Agung Sri Sidhimantra[4], Asmunin[5]
[1*,2,3,4,5]Universitas Negeri Surabaya, Surabaya, Indonesia

ABSTRACT

Keywords:
CO_2 Monitoring System
MQ-135 sensor
Realtime
Linear Regression
IQ Air

Air quality in classrooms is very important for health and comfort. Increased concentrations of carbon dioxide (CO_2) due to human activities can cause health problems such as headaches, fatigue and respiratory problems. This research develops an Internet of Things (IoT) based air quality monitoring system using the HC-SR04 sensor to count the number of people and the MQ-135 sensor to measure CO_2 levels. Data is integrated with the Thinger.io platform in realtime and classified based on IQ Air standards. The IQ Air classification consists of : >1600 PPM in the good category, <=2500 PPM in the moderate category, <=3300 PPM in the unhealthy category for sensitive groups, <=4100 PPM in the unhealthy category, <=4900 PPM in the very unhealthy category, and <=5500 PPM in the hazardous category. The LED indicator and buzzer provide visual and audio warnings based on detected CO_2 levels. The linear regression method was used to calibrate the MQ-135 sensor, showing a high level of accuracy with an error percentage of 1.9%. The test results show that this system provides accurate and realtime information to monitor indoor air quality, provides early warning if CO_2 levels reach unhealthy levels, and can improve air quality in classrooms.

INTRODUCTION

Rooms can have various sizes and functions, and they are used for various purposes according to building needs and design (Work, 2024). In a room there is air which is needed by humans and other living creatures to fulfill human needs (Education, 2024). The maximum capacity of a room should not exceed 50% of its maximum capacity, because this can affect air quality (Certified Commercial Property Inspectors Association, 2024). One important parameter that needs to be monitored is carbon dioxide (CO_2) levels (Meter, 2023). The air can be made less fresh and clean by excessive amounts of carbon dioxide (CO_2) (Meter, 2023).

If there are excessive levels of carbon dioxide (CO_2) in the air, it is also air pollution, which can lead to health problems (WHO, 2022). It states that high levels of carbon dioxide are enough to reduce human cognitive abilities, such as the ability to make basic decisions being reduced by around 25% and the ability to think strategically by 50% (Karnauskas et al., 2020). The average concentration of carbon dioxide (CO_2) in the atmosphere remains below 350 parts per million (PPM) (NASA, 2024). The government has set a safe limit for carbon dioxide (CO_2) in Minister of Health Regulation number 1077 concerning Guidelines for Indoor Air Health, namely 1000 PPM in 8 hours (Ministry of Health, 2023). In most cases, the formation of carbon dioxide (CO_2) occurs due to burning of waste and the use of certain tools, such as air conditioners, which pollute the air because they contain dangerous pollutants (Semarang, 2020).

High levels of carbon dioxide (CO_2) in a room can be caused by the large number of people in the room, the wind emitted from the hot AC, and lack of ventilation (Kumar et al., 2022). Human capacity that exceeds the limits of the room reduces blood flow because the oxygen supply is reduced (Jorner & Casey, 2024). Low blood flow is very dangerous for the brain due to the lack of oxygen being transported through the blood (Harsono, 2020). Carbon dioxide levels that are too high will cause a serious health problem called acidosis (Nandy, 2023).

With current technological developments, monitoring carbon dioxide (CO_2) levels so that they are safe and comfortable can be monitored remotely via Internet of Things (IoT) technology. The Internet of Things (IoT) is used for process automation because it is defined as communication between objects and objects (IBM, 2024). With Internet of Things technology, carbon dioxide in classrooms can be monitored remotely, preventing classroom use that could endanger health. In this research, the process for collecting data was used using the linear regression method. Linear Regression is a statistical method used to create a model that describes how the dependent variable (dependent, response, Y) is related to one or more independent variables (independent, predictor, X) (Wang et al., 2024).

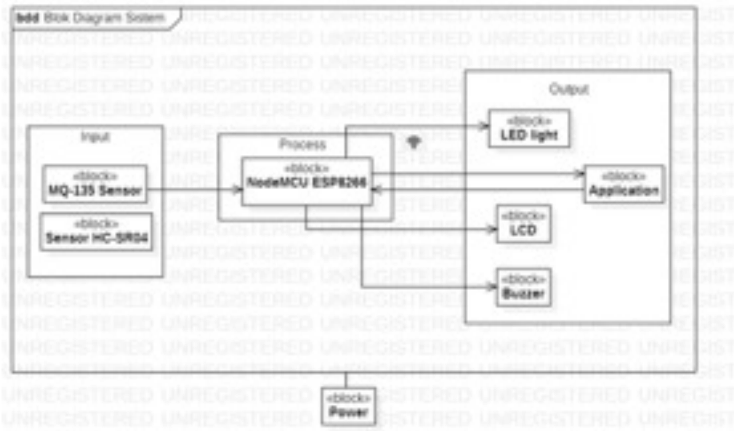

Figure 1. Block Diagram

In Figure 1, the system will use the nodemcu esp8266 as a microcontroller to operate the system. In this research, several inputs were used, such as the MQ-135 sensor for monitoring indoor carbon dioxide (CO_2) levels which produces output in the form of carbon dioxide (CO_2) level values in the form of PPM (Parts Per Million) and the number of people manually. The data read by the sensor is sent to the nodemcu esp8266 microcontroller for processing. Data processed by the microcontroller is forwarded to the system output. The output produced in this research is data obtained by sensors displayed in realtime on the thinger.io website.

RESEARCH METHOD

A method for predicting the value of one variable from another variable. The goal is to estimate the dependent variable (Y) based on the given independent variable (X). The final result of this method is a linear relationship between the independent variable (X) and the dependent variable (Y). Here is the formula:

(1)

In this research, the relationship between CO_2 levels and the number of people present in a room is examined. The CO_2 levels, represented as the dependent variable Y, are influenced by the number of occupants, denoted as the independent variable X. To

understand this relationship, a linear regression model is utilized. In this model, the point where the regression line intersects the Y-axis is known as the intercept, represented by a. This intercept indicates the baseline level of CO_2 when there are no people present in the room. The slope of the regression line, represented by b.
The ratio between the observed value and the predicted value is divided by the observed value and multiplied by 100%. It provides a percentage measure of the relative error between observed and predicted data. The smaller the error percentage, the better the regression model is at predicting the true values. Here is the formula:

(2)

MAPE is used to evaluate the accuracy of prediction models in statistics and machine learning. MAPE provides a measure of the average prediction error as a percentage of the true value, making it easier to interpret how accurate the model is. The lower the MAPE value, the better the model or prediction performance. Here is the formula:

(3)

RESULTS AND DISCUSSION

Data Collection

Data collection was carried out by observation in the Vocational Faculty environment in the K2.04 classroom. The population used in this research was 35 students. The data taken is the number of people in the classroom and carbon dioxide (CO_2) levels. Data for the number of people was taken using the HC-SR04 sensor and data for carbon dioxide (CO_2) levels using the MQ-135 sensor. In table 1, is a data table before the application of the method.

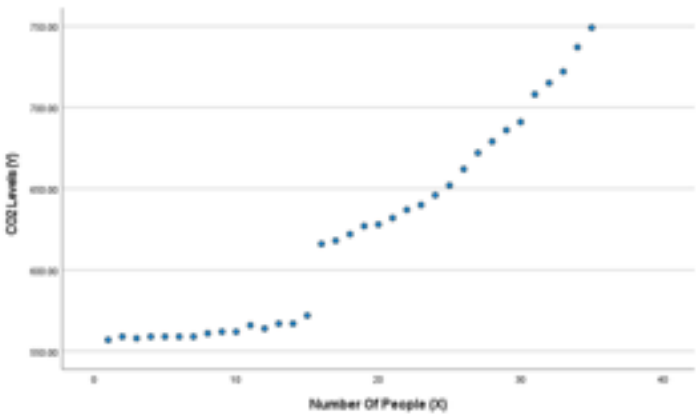

Figure 2. Actual Data Scatter

In Figure 2, shows that in the actual data, there is a very strong positive relationship between the number of people and CO_2 levels, with CO_2 levels increasing as the number of people increases. The high R^2 value (0.9284) shows that this linear regression model is very good at explaining the variability of CO2 levels based on the number of people. This indicates that the measurement of actual CO_2 levels is highly dependent on the number of people. The next process is implementing linear regression to retrieve prediction data by calibrating the sensor. The prediction data obtained will later be used to monitor classroom conditions. With the results a = 5.7692, b = 518.15.

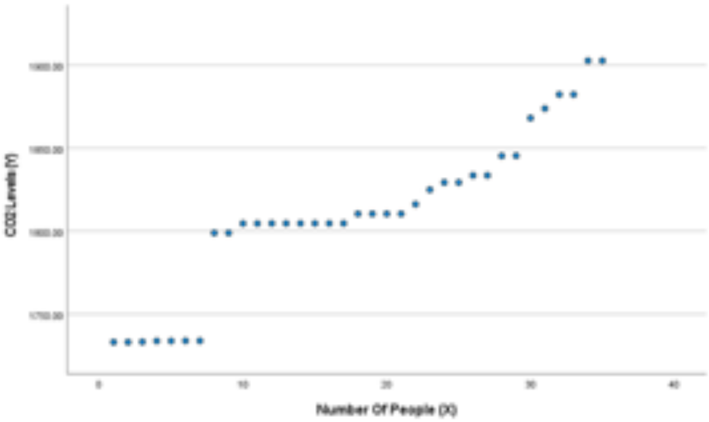

Figure 3. Prediction Data Scatter

In Figure 3, shows that in the predicted data, there is a strong positive relationship between the number of people and CO_2 levels, with CO_2 levels increasing as the number of people increases. The high R^2 value (0.8934) shows that this linear regression model is very good at explaining the variability of CO_2 levels based on the number of people. This indicates that the prediction of CO_2 levels can be highly dependent on the number of people in the model.

Data Analysis

In table 1 is used to provide important information about the independent variables (predictors) being analyzed to determine their effect on the dependent variable (outcome). In table 2 is used to evaluate the performance of the regression model, ensuring that the model is robust and accurately captures the relationship between the dependent and independent variables. In table 3 is used to decompose the variance in the dependent variable into components attributable to the model (regression) and random error (residuals) which can later provide formal statistical test results for the significance of the regression model as a whole. In table 4 is used to interpret the specific contribution of each predictor in the regression model, which allows a detailed understanding of how changes in the predictor affect the dependent variable.

Table 1. SPSS Table

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Number Of People (X)[b]	.	Enter

a. Dependent Variable: CO2 Levels (Y)

b. All requested variables entered.

Table 2. SPSS Output Table with R^2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.945[a]	.893	.890	16.18726

a. Predictors: (Constant), Number Of People (X)

Table 3. Analysis of Variance

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	72459.186	1	72459.186	276.533	<.001[b]
	Residual	8646.909	33	262.028		
	Total	81106.094	34			

a. Dependent Variable: CO2 Levels (Y)

b. Predictors: (Constant), Number Of People (X)

Table 4. SPSS Variable Equation

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1729.822	5.592		309.355	<.001
	Number Of People (X)	4.505	.271	.945	16.629	<.001

a. Dependent Variable: CO2 Levels (Y)

Data Evaluation

Before calculating the average %error, first calculate the %error. After that, calculate the average %error or what is commonly called the Mean Absolute Percentage Error (MAPE)

Which shows that predictions from the linear regression method have a very good level of accuracy. This means that the accuracy level in this system reaches 98.1%.

System Implementation

At the implementation stage the system as a whole is a picture of the system that has been built in accordance with the system design stages. This system was created to be accessed by everyone.

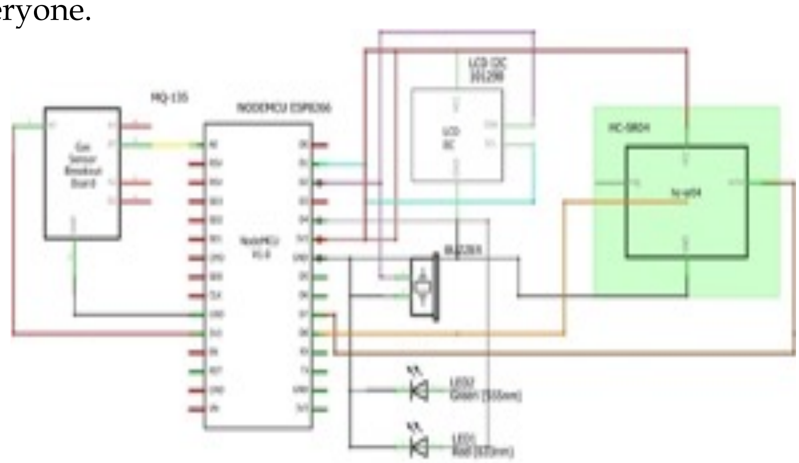

Figure 4. Hardware Schematic Design

In Figure 4, is a prototype of the system. The prototype is designed to meet the requirements required for a carbon dioxide monitoring system. At this stage, an Internet of Things system is implemented consisting of a NodeMCU ESP8266 microcontroller, LED lights, I2C LCD, buzzer, MQ-135 sensor, with operations programmed in C++. In table 5, are the pins of each component attached to the NodeMCU ESP8266.

Table 5. Pin Connection

MQ-135	
Sensor Pin	**NodeMCU ESP8266**
Data (A0)	A0
VCC	3V

GND	GND
HC-SR04	
Sensor Pin	**NodeMCU ESP8266**
Trig	D8
Echo	D7
VCC	3V
GND	GND
LED	
Sensor Pin	**NodeMCU ESP8266**
Plus Pole (+)	D4
Minus Pole (-)	GND
Buzzer	
Sensor Pin	**NodeMCU ESP8266**
Plus Pole (+)	D2
Minus Pole (-)	GND
LCD I2C 16x2	
Sensor Pin	**NodeMCU ESP8266**
SDA	D2
SLC	D1
VCC	3V
GND	GND

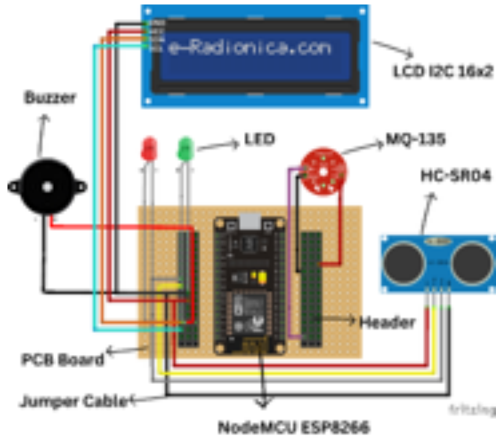

Figure 5. Hardware Design

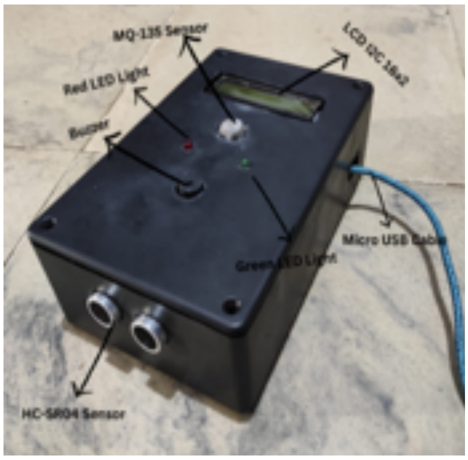

Figure 6. Hardware Packgin Design

In Figure 5, a view of the installed components. In Figure 6, is an image of the hardware packaging where all components have been installed.

Figure 7. Conditions When CO_2 >2500 PPM

Figure 8. Conditions When CO2 <2500 PPM

In Figure 7, when CO_2 > 2500 PPM, the red LED light will light up and the buzzer will turn on. In Figure 8, when CO_2 < 2500 PPM, the green LED light will light up and the buzzer will not sound.

Login Page

Figure 9. Login View

In Figure 9, is something that is done before entering the dashboard of the system, you must log in/create an account first. If you don't have an account then select the create an account menu, if you already have an account then log in directly.

Statistics page

Figure 10. Statistics Display

In Figure 10, after successfully logging in, the initial display of this application is statistics. Statistics is used to provide in-depth insight and analysis regarding device usage and collected data.

Devices Page

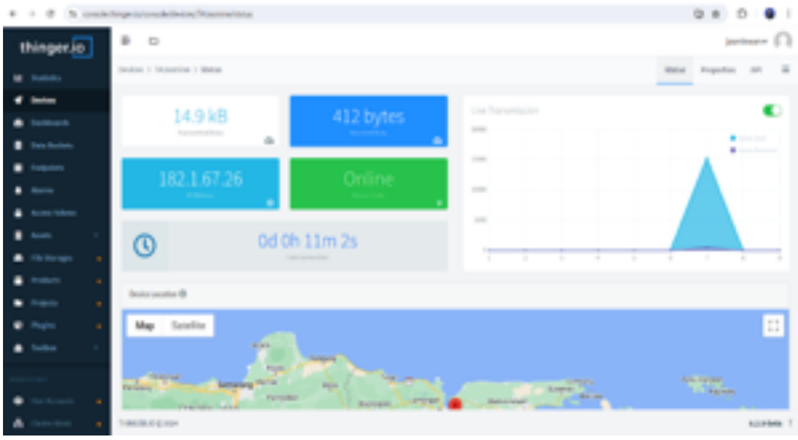

Figure 11. Devices Display

In Figure 11, The devices menu is the management center for devices connected to the platform. Here are some of the main functions and uses of the device menu:

1. Device Management
    a. Addition of new devices.
    b. Device removal.
2. Device Configuration
    a. Device details such as connection status, IP address and last time the device was connected.
    b. Parameter settings such as data capture interval, sensor settings, and network settings.
3. Device Monitoring
    a. Device Connection Status in realtime.

Dashboard Page

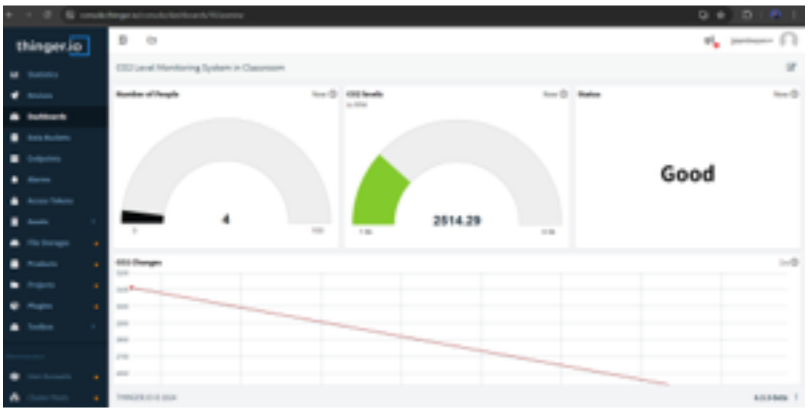

Figure 12. Dashboard View

In Figure 12, the dashboard menu displays features that allow users to create data visualizations and interactive controls for the IoT devices they create. Dashboard is a very useful interface for monitoring and managing data from various devices in realtime.

In this menu you can customize the widget according to your needs. Can display visualization of the data obtained in the form of graphs, charts and gauges. Can control devices with buttons, switches, sliders, and input fields. Can support data integration from various devices and sources, so users can create dashboards that combine data from many sensors or devices. Can set user permissions and share dashboards with teams or the public by setting appropriate access levels.

CONCLUSION

In this research, the HC-SR04 sensor and MQ-135 sensor were successfully integrated into a realtime monitoring system.Implementation of LEDs and buzzers as visual and sound indicators work well. The red LED lights up and the buzzer sounds when CO2 concentrations exceed healthy thresholds, providing a clear warning to users. The implementation of linear regression can provide a percentage error value of 1.9%, which means that predictions from the linear regression method have a very good level of accuracy, namely 98.1%, because the prediction error is relatively small compared to the actual value. The system successfully classifies room conditions based on CO2 concentration. This allows users to know the air quality conditions in realtime and take necessary actions.

REFERENCES

Certified Commercial Property Inspectors Association. (2024). *Commercial Property Safety Requirements: Maximum Occupancy*. Ccpia.Org. https://ccpia.org/occupancy-load-signs/

Education, U. C. for S. (2024). *What's in the Air?* Scied.Ucar.Edu. https://scied.ucar.edu/learning-zone/air-quality/whats-in-the-air

Harsono, F. H. (2020). *Cerebral Hypoxia, Lack of Oxygen Leading to Death*. Www.Liputan6.Com. https://www.liputan6.com/health/read/2690813/hipoksia-serebral-kurang-oksigen-yang-berujung-kematian?page=2

IBM. (2024). *What is the IoT?* Www.Ibm.Com. https://www.ibm.com/topics/internet-of-things

Jorner, M. J., & Casey, D. P. (2024). *Regulation of Increased Blood Flow (Hyperemia) to Muscles During Exercise: A Hierarchy of Competing Physiological Needs*. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4551211/

Karnauskas, K. B., Miller, S. L., & Schapiro, A. C. (2020). Fossil Fuel Combustion Is Driving Indoor CO2 Toward Levels Harmful to Human Cognition. *GeoHealth*, *4*(5). https://doi.org/10.1029/2019GH000237

Kumar, P., Hama, S., Abbass, R. A., Nogueira, T., Brand, V. S., Wu, H. W., Abulude, F. O., Adelodun, A. A., de Fatima Andrade, M., Asfaw, A., Aziz, K. H., Cao, S. J., El-Gendy, A., Indu, G., Kehbila, A. G., Mustafa, F., Muula, A. S., Nahian, S., Nardocci, A. C., … Shiva Nagendra, S. M. (2022). CO2 exposure, ventilation, thermal comfort and health risks in low-income home kitchens of twelve global cities. *Journal of Building Engineering*, *61*(August), 105254. https://doi.org/10.1016/j.jobe.2022.105254

Meter, C. (2023). *High CO2 Levels Indoors Will Surprise You*. Www.Co2meter.Com. https://www.co2meter.com/blogs/news/high-carbon-dioxide-co2-levels-indoors

Ministry of Health. (2023). Minister of Health Regulation No. 2 Year 2023. *Ministry of Health of the Republic of Indonesia*, *55*, 1–175.

Nandy. (2023). *Know the Effects of Carbon Dioxide on Body Health*. Www.Gramedia.Com. https://www.gramedia.com/best-seller/pengaruh-karbondioksida-bagi-kesehatan-tubuh/

NASA. (2024). *Carbon Dioxide*. Climate.Nasa.Gov. https://climate.nasa.gov/vital-signs/carbon-dioxide/?intent=121

Semarang, E. S. (2020). *What Are the Characteristics of Polluted Air?* Dlh.Semarangkota.Go.Id. https://dlh.semarangkota.go.id/seperti-apa-ya-ciri-ciri-udara-yang-tercemar/

Wang, Y., Huang, Q., Yao, Z., & Zhang, Y. (2024). On a class of linear regression methods. *Journal of Complexity*, 101826. https://doi.org/10.1016/j.jco.2024.101826

WHO. (2022). *Ambient (outdoor) air pollution*. Www.Who.Int. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Work, I. (2024). *Rooms With Specific Functions*. Www.Idealwork.Com. https://www.idealwork.com/functionality-of-indoor-environments/