Buletin Poltanesa Vol. 24 No. 2 (December 2023) 235-241 p-ISSN 2721-5350 e-ISSN 2721-5369
https://doi.org/10.51967/tanesa.v24i2.2946 © 2023 Politeknik Pertanian Negeri Samarinda

Air Pollution Assessment of Samarinda Using the C4.5 Algorithm

Anton Prafanto*
Informatic, Mulawarman University, Samarinda
antonprafanto@fkti.unmul.ac.id
**Corresponding author*

Indah Fitri Astuti
Informatic, Mulawarman University, Samarinda
indahfitriastuti@unmul.ac.id

Ummi Salamah
Informatic, Mulawarman University, Samarinda
ummi.syarofhana@gmail.com

Fahrul Agus
Informatic, Mulawarman University, Samarinda
fahrulagus@unmul.ac.id

Awang Harsa Kridalaksana
Informatic, Mulawarman University, Samarinda
awanghk@unmul.ac.id

Vina Zahrotun Kamila
Information System Department, Mulawarman University, Samarinda
vinakamila@ft.unmul.ac.id

Abstract**—** The degradation of air quality in numerous Indonesian cities is attributed to the swift proliferation of motorised vehicles, rapid population growth, and inadequate green spaces. Samarinda, the capital of East Kalimantan province, is plagued by high levels of pollution resulting from heavy vehicle exhaust emissions. The provision of accurate air quality information can mitigate respiratory issues. However, the public does not have access to air quality information due to the high cost of air quality measuring devices. Therefore, an Internet of Things (IoT)-based air pollution monitoring system using ESP32 is needed to provide interactive and real-time information. This study tested the C4.5 algorithm to classify air quality data based on six measurement parameters: PM10, PM2.5, CO, O3, and NO2. PM10 and PM2.5 particles are the primary pollutants that significantly impact human health. The World Health Organization (WHO) has set an annual quality standard value of 20µg/m3 for PM10 and 10µg/m3 for PM2.5. Carbon Monoxide (CO) can reduce the blood's ability to carry oxygen, which can affect the function of vital organs such as the heart and brain. Ozone (O3) on the Earth's surface is a harmful pollutant that can damage the lungs and other respiratory systems. Nitrogen dioxide (NO2) can cause lung inflammation and lower immunity to infections, such as influenza and pneumonia. This study uses the C4.5 algorithm to classify air quality data based on these parameters, which are important for determining air quality. The results show that air quality is divided into two types: good and moderate, with different proportions each day. The C4.5 algorithm achieved a success rate of 99.5074% and a failure rate of 0.4926% when processing air quality data. It was effective in classifying air quality and processing data. An Internet of Things (IoT)-based air pollution monitoring system using ESP32 is needed to provide interactive and real-time information to the public.

Keywords**—**ESP32, Air quality, C4.5 Algorithm, Internet of Things, Classifying air quality.

I. Introduction

Air pollution occurs when substances, energy, and other components enter the air as a result of human activities, causing air quality to drop to an unhealthy level that can harm human health (Decree of the Indonesian Minister of Health Number 1407/MENKES/SK/XI/2002). Clean air quality must be prioritised, as low air quality can directly impact the health of living organisms on the Earth, particularly humans (Rahmawati & Khairina, 2021). Some of the health consequences resulting from exposure to air pollution are headaches, dizziness, nausea, irritation of the nose, throat, eyes and skin, acute respiratory infections (ARI), asthma, and the deadliest disease, pneumonia (Sari & Mayasari, 2020).

The high level of human activity and population density in urban areas has resulted in significantly higher levels of air pollutants compared to other areas (Rahmawati & Khairina, 2021). The degradation of air quality in multiple cities in the Indonesian region persists in its escalation (Greenstone & Fan, 2019). This phenomenon has multiple causes. One of them is the swift progress of motorised transportation, which is directly linked to the increase in population, but not correlated with the expansion of green open spaces and the conservation of green zones, particularly in urban regions.

Samarinda, the capital city of East Kalimantan province in Indonesia, is the largest city on the island of Kalimantan with a population of 858,080 residents, consisting of 443,379 males and 414,701 females (Central Bureau of Statistics, 2019). The city spans an area of 718 km², possessing hilly terrain with altitudes ranging from 10 to 200 metres above sea level. Excessive amounts of exhaust emissions from motorised vehicles contribute to the high levels of pollution experienced in the city. Every year, there has been a noticeable increase in the use of motorised vehicles in Samarinda City, as stated by the Central Bureau of Statistics (BPS) of East Kalimantan Province (Hidayat & Noor, 2020).

Following the various challenges outlined above, a proposal was put forward to establish an air pollution monitoring system aimed at categorising air quality in a city through the application of ISPU using the C4.5 algorithm. The algorithm in question was created by Ross Quinlan in 1993 and represents a modification of the ID3 algorithm. In light of its distinct features, the C4.5 algorithm is considered to offer significant advantages over similar decision tree algorithms, particularly in its ability to develop classification models that deliver high levels of accuracy, even when dealing with complex data. The C4.5 algorithm creates a classification model that is straightforward to interpret, enabling comprehension of the relationship between input variables and output variables. Moreover, the C4.5 algorithm is relatively efficient in terms of computation time, making it suitable for larger datasets (Feng & Zhang, 2023).

In addition to the C4.5 algorithm, other algorithms are available to classify air quality based on ISPU. These include the Naive Bayes, K-Nearest Neighbors, and Support Vector Machine algorithms, each with its own set of strengths and weaknesses. Although the Naive Bayes algorithm computes quickly, its resulting accuracy is relatively subpar. The K-Nearest Neighbors algorithm offers higher accuracy, but it presents interpretational challenges. On the other hand, the Support Vector Machine algorithm provides accuracy and interpretation advantages, but its computation speed is relatively slow (Damayunita et al., 2022).

The ISPU data utilized in this study resulted from monitoring using an ESP32 device that has an array of sensitive gas sensors. This device has demonstrated its usefulness in air quality monitoring since it can measure a variety of pertinent gas parameters. The gas sensors measure carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), and various particulates. To incorporate the data from the sensors, the ESP32 device is connected to Thinger.io's IoT platform (Dobrzyniewski et al., 2022).

Thinger.io is an Internet of Things platform that facilitates the connection of physical devices to the internet and the transmission of data to the platform. The platform enables the seamless integration of various sensors and devices with Thinger.io, which is particularly advantageous in the context of this research, where integration and processing of data from different gas sensors is necessary (Yuan et al., 2023). Moreover, the platform allows for real-time transmission of sensor data and provides features for further data analysis.

The information gathered from ESP32 devices and transmitted to Thinger.io is subsequently exported in a data analysis software-compatible format. For this research project, the data was exported in Excel format, which is commonly used in data analysis. Excel provides researchers with a convenient way to handle and arrange data in preparation for further analysis.

The data were analyzed using the widely used open-source Weka software for data modeling and analysis. Weka offers various classification algorithms, including the C4.5 algorithm employed in this research. Weka's flexibility in data processing and analysis enables researchers to conduct diverse experiments to assess the performance of different classification algorithms.

Utilizing ESP32 hardware, Thinger.io IoT platform, and Weka analysis software, this study seeks to gain a comprehensive comprehension of air quality based on ISPU. The ESP32 data encapsulates numerous pertinent factors, and diligent scrutiny of this data can unveil correlations between various air pollution agents. The research findings offer insights into improving air quality monitoring and efforts towards maintaining clean and healthy air. Additionally, they can be used to create more effective policies aimed at achieving these objectives.

II. Methods

A. Air Quality Based on Air Pollutant Standard Index (ISPU)

Air pollution causes changes in the air environment by introducing pollutants, including gases and small particles or aerosols. These pollutants enter the air through natural means, such as forest fire smoke, volcanic ash, meteorite dust, and salt emissions from the sea, as well as through human activities such as transportation and industrial waste disposal (Syihabuddin Azmil Umri, 2021). The high concentration of air pollution in various major cities and industrial regions in Indonesia leads to respiratory problems, eye and ear irritation, the development of specific diseases, and impaired visibility (Abidin & Artauli Hasibuan, 2019). Judging by physical characteristics, air pollutants can be:

a. Particles (dust, aerosols, lead)
b. Gas (CO, NO2, SO2, H2S, and HC)
c. Energy (temperature and noise)

On the basis of their occurrence, pollutants are formed by:

a. Primary pollutants (which are emitted directly from the source)
b. Secondary pollutants (which are formed as a result of reactions in the air between various substances)

Air quality standards are measures of the permissible levels of air pollutants that can exist in the atmosphere. Ambient air refers to the free air in the troposphere that is essential for human health, living organisms, and the environment. To prevent air pollution, the maximum limit of air quality in ambient air is established as the national ambient air quality standard, as detailed in PP No. 41 of 1999.

The quality standard value unit in Indonesia employs mass units per volume represented as µg/Nm3. In the United States, in addition to using mass per unit of volume, the ratio of volume per unit of volume is also utilised, denoted as ppm (parts per million), which indicates the amount of pollutants in cm3 per m3 of air, as well as ppb (parts per billion). The standard value for each pollutant parameter varies between different countries. This is adjusted to fit the environmental conditions within each respective country. The lower the standard value, the greater the harm to environmental health. Tables 1, 2, and 3 demonstrate that air pollution can be measured in parts per million (ppm), representing the amount of pollutant in cubic centimetres per cubic meter of air, or in micrograms per cubic meter.

Table 1. National Ambient Air Quality Standard

No.	ISPU VALUE	Measurement Time	Quality Standard
1	SO_2 (sulphur dioxide)	1 Hour	900 ug/Nm^3
		24 hours	365 ug/Nm^3
		1 year	60 ug/Nm^3
2	CO (carbon monoxide)	1 Hour	30.000 ug/Nm^3
		24 hours	10.000 ug/Nm^3
		1 year	-

Table 2. Basic Parameters For Air Pollutant Standard Index (ISPU) And Measurement Time Period

No.	Parameter	Average Measurement Time
1.	Particulate matter (PM10)	24 hours
2.	Sulphur Dioxide (SO2)	24 hours
3.	Carbon monoxide (CO)	8 hours
4.	Ozone (O3)	1 hour
5.	Nitrogen Dioxide (NO2)	1 hour

Table 3. Air Pollutant Standard Index Limit (In Units Of Si)

Air Pollutant Standard Index	24 hours PM10 ug/m3	8 Hours SO2 ug/m3	8 Hours CO ug/m3	1 hour O3 ug/m3	1 hour NO2 ug/m3
50	50	80	5	120	
100	150	365	10	253	
200	350	800	17	400	1130
300	420	1600	34	800	2260
400	500	2100	46	1000	3000
500	600	2620	57.5	1200	3750

The air quality index value is calculated using the following equation (United States Environmental Protection Agency, 2018):

$$I = \frac{(Ia-Ib)}{(Xa-Xb)}\,(Xx - Xb) + Ib \quad (1)$$

Information : (1)
I : calculated ISPU
Ia: ISPU upper limit
Ib : ISPU lower limit
Xa : ambient level (µg/m3)
Xb : lower limit ambient level (µg/m3)
Xx : real ambient level of measurement result (µg/m3)

The known ambient air concentration for the parameter type SO2 is: 322 µg/m3. The concentration if converted into the air pollutant standard index number is as follows:

Xx : Real ambient levels of measurement results: 322 µg/m3
Ia : ISPU upper limit: 100 (line 2)
Ib : ISPU lower limit: 50 (line 3)
Xa : Ambient upper limit: 365 (line 4)
Xb : Ambient lower limit: 80 (line 5)

So these numbers are entered into formula (*) to become:

$$I = \frac{(Ia - Ib)}{(Xa - Xb)}\,(Xx - Xb) + Ib$$

$$I = \frac{(100 - 50)}{(365 - 80)}\,(322 - 80) + 50$$

The SO2 concentration of 322 mg/m3 in the ambient air is converted to the Air Pollution Standard Index (ISPU), resulting in a value of 92. The SO2 concentration of 322 mg/m3 in the ambient air is converted to the Air Pollution Standard Index (ISPU), resulting in a value of 92. This value is utilised to classify air quality in a location. Based on the five ISPU categories, each is represented by a colour symbol. The good category is green, the moderate category is blue, the unhealthy category is yellow, the very unhealthy category is red, and the dangerous category is black.

Each pollutant used as an ISPU parameter, as determined by the Decree of the Head of the Environmental Impact Control Agency, has an assigned impact level ranging from good to dangerous. Table 4 provides a summary of the effects of ISPU for each parameter (Agista et al., 2020).

Table 4. Effect of air pollutant standard index for each pollutant parameter

Category	Range	CO	SO2
Good	0-50	No effect	Wounding of certain plant species when combined with O3 exposure for a duration of 4 hours.
Medium	51 - 100	No changes in blood chemistry were detected	Injury to some plant species
Unhealthy	101 - 199	Cardiovascular improvement in smokers with heart disease	Increased Plant Damage from Odor
Very Unhealthy	200-299	Cardiovascular improvement in non-smokers with heart disease, and there will be some noticeable weaknesses	Increased sensitivity in patients with asthma and bronchitis
Harmful	300 - more	Levels harmful to all populations when exposed	

B. *Algorithm C4.5*

The C4.5 algorithm is commonly used to classify segmentation or predictive grouping through the use of decision trees to aid decision making. Its notable advantages include processing continuous and discrete numeric data, the ability to handle missing attribute values, the production of easily interpretable rules, and being one of the fastest algorithms available (Marlina & Bakri, 2021). Prediction accuracy refers to the ability of the model to perform well in predicting class labels for new or previously unknown data. In terms of the computational speed or efficiency necessary to build and employ the model (Febriani & Sulistiani, 2021), as well as the model's capability to make accurate predictions even in the presence of missing attribute values.

C. Hardware Design

Block diagram depicting the circuitry and indicating the pins that are connected, along with a detailed description of the communication protocol utilized for this study.

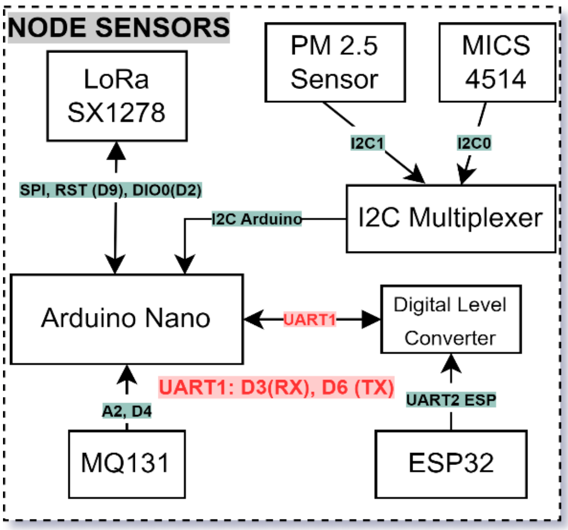

Figure. 1. Circuit Block Diagram

1) Arduino Nano

The Arduino Nano serves as the central hardware component of the air quality and weather monitoring system. It receives data from multiple sensors and LoRa modules, which can then be transmitted to the Gateway for analysis.

2) PM 2.5

The PM 2.5 sensor detects particles in the air with a diameter of less than 2.5 micrometres. Such small particles can reduce air quality and pose health risks.

3) PM 2.5

The MICS 4514 gas sensor detects exhaust emissions, including CO, NO2, and SO2, which have significant negative impacts on air quality. The MICS 4514 gas sensor detects exhaust emissions, including CO, NO2, and SO2, which have significant negative impacts on air quality. This objective statement and its concise structure avoid biased language while maintaining a precise and logical flow of information.

4) MQ-131

The MQ-131 gas sensor detects concentrations of ozone gas, which can irritate the eyes and respiratory system. The MQ-131 gas sensor detects concentrations of ozone gas, which can irritate the eyes and respiratory system. The MQ-131 gas sensor detects concentrations of ozone gas, which can irritate the eyes and respiratory system. Its measurement of ozone concentration is vital to determining exposure levels.

An IoT platform connected to an integrated air quality monitoring system enables quick and easy access to information on air conditions by users. This can help the public and associated organisations make more informed decisions when addressing the negative consequences of inadequate air quality.

III. Results and Discussion

A. Application/Data Processing

In this study, the focus is on monitoring air quality using a set of prototype devices, with subsequent processing of data obtained from IoT instruments. The aim is to determine the level of air quality.

1) Data Preparation/Pre-Processing

Sensor data used in this research were obtained from the Antares website in CSV format. The Antares website serves as a storage platform for the sensor data of this device. The dataset comprises multiple individual files for each month from March to August 2023. An explanation of the variables or parameters contained in the data set is as follows:

a. Date: time of air quality measurement.
b. PM10: air particles smaller than 10 micrometres (micrometres).
c. PM2.5: airborne particles smaller than 2.5 micrometres (micrometres).
d. CO: Carbon monoxide.
e. O3: a radical inorganic molecule consisting of three oxygen atoms that is a strong oxidiser. Ozone is naturally produced from oxygen (O2) molecules in the atmosphere of the Earth that interact with ultraviolet light or electrical activity in the atmosphere.
f. NO2: Nitrogen dioxide.
g. Max: The highest measured value of all the parameters measured at the same time.
h. Critical: the parameter with the highest measurement result.
i. Category: obtained from the average calculation of PM10. PM25, SO2, CO, O3, and NO2, then classified according to ISPU (Air Pollution Standard Index).

Table 5 displays the National Ambient Air Quality Standards, explaining the categories, corresponding colors, and air quality values to enhance the public's comprehension.

Table 5. National Ambient Air Quality Standard

Category	Status Colour	Number range
Good	Green	1 – 50
Medium	Blue	51 – 100
Unhealthy	Yellow	101 – 200

Very unhealthy	Red	201 - 300
Dangerous	Black	⩾ 301

Overall, the dataset comprises 8,760 entries with 10 variables and one class. The variables of the dataset include date, time, PM10, PM2.5, SO2, CO, O3, NO2, Max, Critical Component, and Category as classes.

After the data has been cleaned, it is now ready for processing. Table 7 displays the processed data available for classification through the C4.5 Algorithm using the Weka tool.

Table 6. Processed Sample Data

Date	Time	PM10	PM2.5	CO	O3	NO2	Max	Critical Component	Category
05/03/2023	02:00:00	25	56	7	0	6	56	PM2.5	Moderate
	03:00:00	25	56	7	0	6	56		Moderate
	04:00:00	25	56	7	0	6	56		Moderate
	05:00:00	25	56	7	0	6	56		Moderate
	06:00:00	26	57	7	0	6	57		Moderate
	07:00:00	26	57	7	0	6	57		Moderate
	08:00:00	27	58	7	0	6	58		Moderate
	09:00:00	28	59	8	0	6	59		Moderate
	10:00:00	28	59	8	0	6	59		Moderate
	11:00:00	29	60	8	0	7	60		Moderate
	12:00:00	29	60	8	0	7	60		Moderate
	13:00:00	29	60	8	0	7	60		Moderate
	14:00:00	30	61	9	0	7	61		Moderate
	15:00:00	30	61	9	0	7	61		Moderate
	16:00:00	30	61	9	0	7	61		Moderate

B. Application/Data Processing

The C4.5 algorithm implements a decision tree by initially choosing an attribute as the root. From there, a branch is generated for every value in the root. The following step is to divide the cases into branches. This process is repeated for each branch until all cases in the branch are of the same class. During root attribute selection, the highest gain value of the available attributes is used as a basis. To compute the gain value, we apply the following formula (IBM, n.d.):

$$Gain\ (S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(S) \tag{2}$$

Information : (2)
S : The set of cases
A : Attributes
N : Number of partitions of attribute A
|Si| : Number of cases in the i-th partition
|S| : Number of cases in S-Meanwhile, the calculation of the entropy value can be seen in the equation :

$$Entropy\ (S) = \sum_{i=1}^{n} - \mathrm{pi} * \mathrm{log2\ pi}$$

Figures 2 and 3 the outcomes of Weka tool's classification and decision tree calculations.

```
Classifier output

Number of Leaves  :     45

Size of the tree :      89


Time taken to build model: 0.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        8485               99.5074 %
Incorrectly Classified Instances        42                0.4926 %
Kappa statistic                          0.9886
Mean absolute error                      0.0064
Root mean squared error                  0.0681
Relative absolute error                  1.4749 %
Root relative squared error             14.6567 %
Total Number of Instances             8527

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.996    0.007    0.997      0.996    0.996      0.989    0.997     0.998     Good
                 0.993    0.004    0.992      0.993    0.992      0.989    0.997     0.994     Moderate
Weighted Avg.    0.995    0.006    0.995      0.995    0.995      0.989    0.997     0.997

=== Confusion Matrix ===

    a    b   <-- classified as
 5813   22 |    a = Good
   20 2672 |    b = Moderate
```

Figure. 2. Classification output

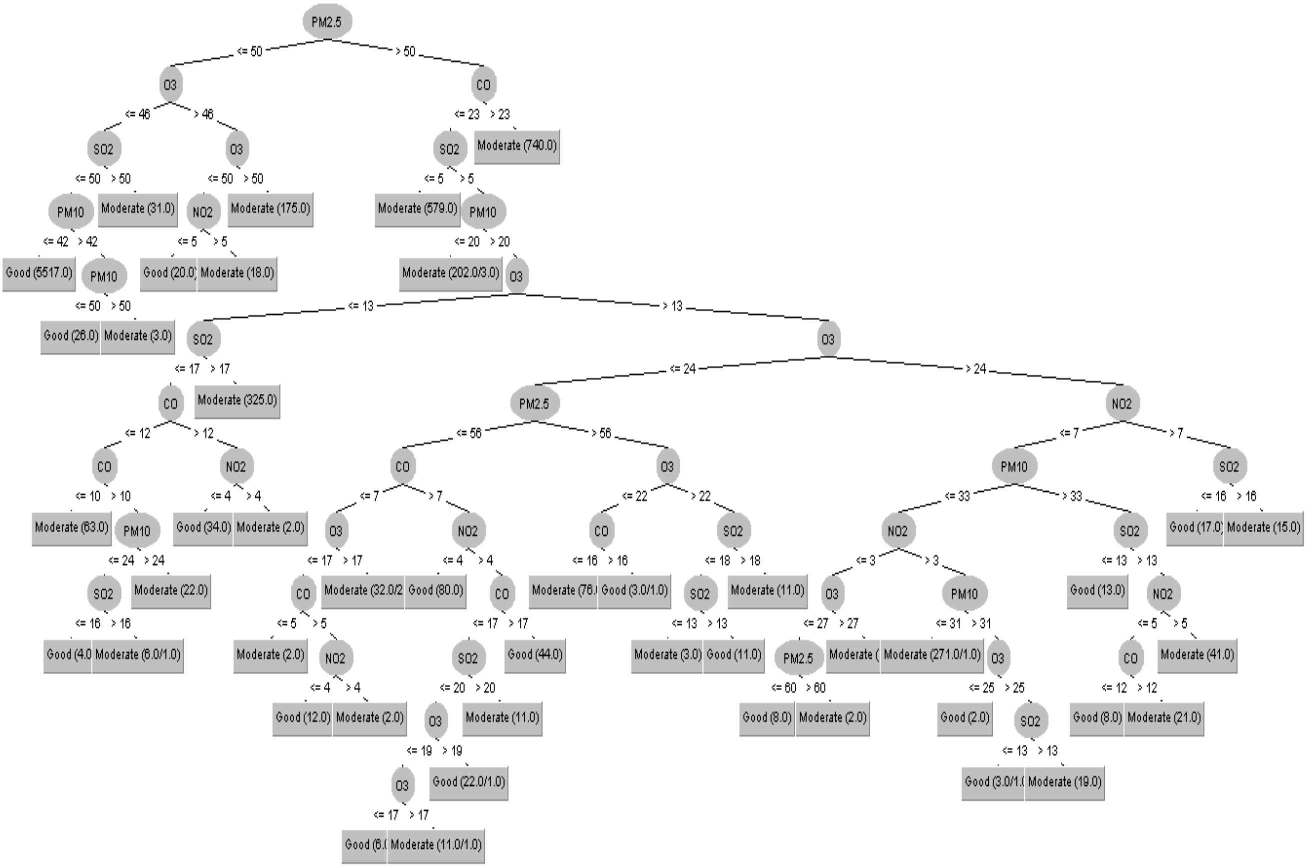

Figure. 3. Tree-shaped classification output

The study produced ten decisions with a success rate of 99.5074% and a failure rate of 0.4926%. The decision tree based on the C4.5 algorithm and presented in Figures 2 and 3 provides the following results. If the PM2.5 value of PM2.5 is above 50 and the amount of CO is higher than 23, then it will be placed in the moderate classification. Lastly, if the PM2.5 value is higher than 50, but the amount of CO is less than or equal to 23, the SO2 level is greater than 5, and the PM10 level is not more than 20, it is also categorized as moderate. If the PM2.5 value is greater than 50, but the amount of CO is less than or equal to 23, and the level of SO2 is not greater than 5, it also falls under the moderate classification. Lastly, if the PM2.5 value is higher than 50, but the amount of CO is less than or equal to 23, the SO2 level is greater than 5, and the PM10 level is not more than 20, it is also categorized as moderate. Lastly, if the PM2.5 value is higher than 50, but the amount of CO is less than or equal to 23, the SO2 level is greater than 5, and the PM10 level is not more than 20, it is also categorized as moderate. If the PM2.5 value is 50 or below and the O3 value exceeds 50, air quality falls into the moderate category.

If the PM2.5 value is 50 or less and the O3 value is 50 or less, while the NO2 value exceeds 5, the air quality falls under the moderate category. If the PM2.5 value is equal to or less than 50 and O3 is equal to or less than 46, and SO2 exceeds 50, it falls into the moderate category. If the PM2.5 value is equal to or less than 50, O3 is equal to or less than 46, SO2 is equal to or less than 50, and PM10 is equal to or less than 42, it falls into the good category. If PM2.5 has a value that is less than or equal to 50 and O3 has a value that is less than or equal to 46 and SO2 has a value that is less than or equal to 50 and PM10 has a value greater than 42 but less than or equal to 50, then it falls into the good category. If PM2.5 has a value that is less than or equal to 50 and O3 has a value that is less than or equal to 46 and SO2 has a value that is less than or equal to 50 and PM10 has a value greater than 42 and greater than 50, then it falls into the moderate category.

IV. Conclusion

After data analysis and discussion, the C4.5 algorithm was tested to classify air quality data based on five air quality measurement parameters: PM10, PM25, CO, O3, NO2 and category. The results demonstrate a clear classification of air quality using the C4.5 data mining algorithm. The analysis categorises air quality into two types: good and moderate, with varying proportions each day. It is concluded that the air quality between March and August 2023 is quite good and not harmful to living organisms.

The C4.5 algorithm was applied to process air quality data, resulting in a 99.5074% accuracy rate for successful classification and 0.4926% for failed classification. These results demonstrate the effectiveness of the C4.5 algorithm in data processing.

REFERENCES

Abidin, J., & Artauli Hasibuan, F. (2019). Pengaruh Dampak Pencemaran Udara Terhadap Kesehatan Untuk Menambah Pemahaman Masyarakat Awam Tentang Bahaya Dari Polusi Udara. *Prosiding Seminar Nasional Fisika Universitas Riau IV (SNFUR-4)*, *September*, 1–7.

Agista, P., Gusdini, N., & Maharani, M. (2020). Analisis Kualitas Udara Dengan Indeks Standar Pencemar Udara (Ispu) Dan Sebaran Kadar Polutannya Di Provinsi Dki Jakarta. *Sustainable Environmental and Optimizing Industry Journal*, *2*(2), 39–57. https://doi.org/10.36441/seoi.v2i2.491

Damayunita, A., Fuadi, R. S., & Juliane, C. (2022). Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients. *Jurnal Online Informatika*, *7*(2), 219–225. https://doi.org/10.15575/join.v7i2.919

Dobrzyniewski, D., Szulczyński, B., & Gębicki, J. (2022). Application of a Gas Sensor Array to Effectiveness Monitoring of Air Contaminated with Toluene Vapors Absorption Process. *Journal of Ecological Engineering*, *23*(10), 269–282. https://doi.org/10.12911/22998993/152427

Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *89Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *2*(4), 89–95.

Feng, H., & Zhang, X. (2023). A novel encoder-decoder model based on Autoformer for air quality index prediction. *PLoS ONE*, *18*(4 APRIL), 1–18. https://doi.org/10.1371/journal.pone.0284293

Greenstone, M., & Fan, Q. (Claire). (2019). Kualitas udara Indonesia yang memburuk dan dampaknya terhadap harapan hidup. *Air Quality Life Index*, 1–10. https://aqli.epic.uchicago.edu/wp-content/uploads/2019/03/Indonesia.Indonesian.pdf

Hidayat, M. A., & Noor, A. (2020). Pengaruh Pertumbuhan Ekonomi Terhadap Alih Fungsi Lahan di Kota Samarinda. *Inovasi*, *16*(2), 10. http://journal.feb.unmul.ac.id/index.php/INOVASI/article/view/8256

IBM. (n.d.). What is a Decision Tree | IBM. In *www.ibm.com*. https://www.ibm.com/topics/decision-trees

Marlina, D., & Bakri, M. (2021). Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *2*(1), 23–28.

Rahmawati, D. S., & Khairina, R. L. (2021). Pengaruh Kualitas Udara Dalam Ruangan Bagi Performa Akademik Pelajar: Sebuah Tinjauan Literatur. *Js (Jurnal Sekolah)*, *5*(1), 34. https://doi.org/10.24114/js.v5i1.22703

Sari, R. P., & Mayasari, D. (2020). Penatalaksanaan Holistik Penyakit Paru Obstruktif Kronik pada Lansia dengan Riwayat Merokok dan Paparan Polusi Udara. *Medula*, *10*(2), 257–266.

Syihabuddin Azmil Umri, S. (2021). Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta. *JIKO (Jurnal Informatika Dan Komputer)*, *4*(2), 98–104. https://doi.org/10.33387/jiko.v4i2.2871

United States Environmental Protection Agency. (2018). Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). *Environmental Protection*, 22. https://airnowtest.epa.gov/sites/default/files/2018-05/aqi-technical-assistance-document-may2016.pdf

Yuan, L., Qiang, L., Di, C., & Weizhi, L. (2023). Research on the strategy of locating abnormal data in internet of things management platform based on improved modified particle swarm optimization convolutional neural network algorithm. *The Journal of Engineering*, *2023*(4). https://doi.org/10.1049/tje2.12263