Examining Changes in Air Quality Levels Pre and During COVID-19 in DKI Jakarta Using Pearson Correlation

Alma Miskha Azizaha) and Bakti Amirul Jabarb)

Department of Computer Science, BINUS University, Jakarta, Indonesia

a) Corresponding author: alma.azizah@binus.ac.id

b) bakti.jabar@binus.ac.id

**Abstract.** Air pollution refers to the presence of harmful or excessive quantities of substances in the air that pose risks to human health, the environment, and the climate. These pollutants can originate from natural sources or human activities, and they can have immediate and long-term health effects. This research investigates whether reduced human activity during the COVID-19 pandemic improved air quality in Indonesia, specifically examining the correlation between vehicle mobility, rainfall intensity, and air pollution levels. Analyzing data from quarter one 2018 to quarter four 2020, the study finds that a 38.02% decrease in vehicle mobility led to a 55.07% reduction in days with poor air quality. Using Pearson correlation analysis method, it reveals a positive correlation coefficient of 0.401 between vehicle numbers and poor air quality, a strong correlation of 0.734 between higher rainfall and improved air quality and identifies ozone (O3) as a dominant pollutant that contribute to poor air quality with 0.983 correlation coefficient and can cause Acute Respiratory Infection (ARI) with a 0.365 correlation coefficient.

**Keywords:** Poor air quality, COVID-19 Pandemic, Pearson Correlation, Rainfall Intensity, Air Pollution

# INTRODUCTION

Air pollution is a serious global issue that has a definite impact on human health and ecosystems. The act of introducing physical, chemical, or material into the regular air environment to the point where they are detectable is the definition of air pollution [1]. Air pollution can cause climate change through the release of greenhouse gases and certain particles into the atmosphere. If air pollutants exceed the predetermined threshold, it can cause a global-scale greenhouse effect phenomenon. Due to the large levels of air pollutants in the air, sunlight shining on the Earth's surface can be reflected back onto the Earth's surface and cannot be reflected into outer space [2]. If this trend continues, it will contribute to global warming, which in turn will drive significant climate change.

Climate change itself can cause various impacts involving various sectors, including the environment, economy, human health and community life. The clearly visible impacts of climate change are significant changes in local weather and climate patterns, including changes in rainfall patterns and an increase in the frequency of extreme weather [3]. If climate change causes erratic rainfall, air pollution is likely to increase. Rain plays a crucial role in reducing air pollutants by washing them out of the atmosphere. Without regular rainfall, pollutants such as particulate matter and gases can accumulate, leading to higher levels of air pollution and associated health risks. Thus, consistent rainfall patterns are essential for maintaining air quality. Climate change can also cause rising sea levels due to melting ice and thermal expansion of sea water threaten coastal areas and low-lying islands. Ecosystem damage and loss of natural habitat occurs due to changes in temperature and species distribution. In addition, drought and water scarcity are serious problems in some regions, affecting food production, agriculture and clean water supplies [4]. These impacts not only harm the environment and health, but also have significant economic consequences, creating instability in various sectors [5].

The impact of air pollution on human life is profound, and one of the key factors worsening air quality is excessive vehicle mobility. Vehicles significantly contribute to air pollution. According to the Union of Concerned Scientists nis, passenger vehicles and heavy-duty trucks are major sources of harmful pollutants. These emissions degrade air quality, leading to serious health issues such as respiratory illnesses, cardiovascular diseases, and cancer. The transportation sector is responsible for a large proportion of smog-forming emissions in the air. Vehicle emissions play in contributing to air pollution, particularly in the creation of harmful ozone. Specific chemicals emitted from vehicles such as ethylene, propylene, toluene, m/p-xylenes, o-xylene, and propane were identified as the primary contributors to the formation of ozone. These chemicals, known as ozone precursors, react in the atmosphere under sunlight to form ground-level ozone, a key component of smog [1].

However, the emergence of the COVID-19 pandemic has drastically affected human lifestyles. This pandemic has caused changes in social and economic activities, including the implementation of social isolation which limits human mobility and activities. This study specifically focuses on understanding how reduced human activity due to social isolation, particularly changes in vehicle mobility could contribute to improved air quality. By comparing air quality data from the pre-pandemic period with the period when social isolation was implemented and considering various air pollutant parameters such as rainfall intensity, this research seeks to understand the correlation between reduced human activity and changes in air pollution levels and identify the causes and effects of air pollution using pearson correlation.

Pearson correlation perhaps the most well-known correlation to measure the linear relationship between two variables [6]. Because the variables of this study suggests a potential linear relationship, such as an increase in vehicle numbers is expected to lead to a similar increase in pollution levels, Pearson correlation method is implemented in this study.

# METHODS

The method used in this research is data mining, which enabled a comprehensive analysis of air quality trends before and during the COVID-19 pandemic. Data mining itself is the process of collecting and processing data aimed at extracting important information. The data mining process consists of data collection, data extraction, data analysis, and data statistics. These four processes are also known as knowledge discovery, knowledge extraction, data/pattern analysis, and information harvesting.

These four processes in data mining will produce highly useful information [7]. The methods of data development include knowledge discovery, which is one of the four processes of data mining. Knowledge discovery is a method to obtain knowledge from data that has interconnected or related tables. The knowledge discovery process also involves several stages, including Data Selection, Pre-processing, and Transformation [6, 8].

The data collection method involves extracting information from multiple sources, including data.jakarta.go.id and BPS. Data extraction is carried out to obtain certain information [6, 7]. In this study, the author extracted data on DKI Jakarta Air Quality Index and Epidemical Data in Public Health Center of DKI Jakarta from data.jakarta.go.id*, as shown in* ***TABLE 1***. The DKI Jakarta AQI was chosen because it provides a comprehensive measure of air quality, capturing various pollutants that are known to affect respiratory health. The epidemiological data from Public Health Centers were selected to analyze the incidence of health issues, such as Acute Respiratory Infections (ARI), that may be linked to fluctuations in air quality.

**TABLE 1**. Sample of Air Quality Raw Data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Quarter | Date | AQI Station | pm10 | so2 | co | o3 | no2 | max | critical | category |
| 1 | 1/1/18 | DKI1 (Bunderan HI) | 35 | 20 | 22 | 27 | 2 | 35 | PM10 | GOOD |
| 2 | 4/10/19 | DKI2 (Kelapa Gading) | 76 | 19 | 23 | 103 | 18 | 103 | O3 | UNHEALTHY |
| 3 | 7/20/20 | DKI5 (Kebon Jeruk) Jakarta Barat | 60 | 14 | 13 | 119 | 5 | 119 | O3 | UNHEALTHY |
| 4 | 11/16/18 | DKI2 (Kelapa Gading) | 54 | 31 | 12 | 234 | 16 | 234 | O3 | VERY UNHEALTHY |

After all the data has been extracted from the sources, the next stage is data analysis. Data analysis involves transforming data into a desired form to gain deeper insights [9]. Based on the analysis results, the author identified the need for data transformation. Specifically, the categorical data as in **TABLE 1** must be converted into numerical data to enable statistical testing using pearson correlation test. In **TABLE 2**, the "Good Air Quality" category corresponds to the "GOOD" classification in the raw data, while "Poor Air Quality" includes both "UNHEALTHY" and "VERY UNHEALTHY" categories.

**TABLE 2**. Converted Numerical Data

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Quarter | Good Air Quality | Poor Air Quality |
| 2020 | 1 | 33 | 8 |
| 2020 | 2 | 13 | 4 |
| 2020 | 3 | 3 | 11 |
| 2020 | 4 | 28 | 8 |

**The research analysis also identified other additional variables that may significantly contribute to changes in air quality, such as Rainfall Intensity and Vehicle Mobility. These factors were used prove rainfall affects the substances that generate pollution and whether emissions from vehicles have a direct impact on pollution levels. Based on the analysis of Rainfall Intensity data, it revealed that there is a missing value and requires data transformation to address the gap. To resolve this, author reprocessed the data by calculating the average rainfall intensity, aggregating data from the same quarter across multiple years *as seen in* ***FIGURE 1***. This approach restored the dataset's completeness, enabling more accurate statistical analysis.

**FIGURE 1**. Data Pre-processing Illustration

After all the data has been analyzed and transformed, the next stage is data statistics. In this stage, descriptive statistics are employed to summarize and present the data results [10]. All data results that have been analyzed and transformed into the desired form of information are presented in the form of visual graphics. While inductive statistics or also known as inferential statistics, are employed to perform correlation tests [10]. This research utilizes the Pearson Correlation coefficient, facilitated by the scipy.stats library, to conduct the statistical analysis. Pearson correlation, also known as Pearson correlation coefficient as in (1), is a method of statistical analysis used to determine the relationship between two continuous variables [9, 11].

A black math symbols with a line

Description automatically generated with medium confidence

(1)

In this analysis, the Pearson correlation coefficient (𝑟) is used to determine the degree of relationship between the variables analyzed. This coefficient can be a positive, negative, or zero value, indicating the direction and intensity of the relationship between these variables [11, 12].The Pearson correlation coefficient, often denoted as 𝑟, where 𝑟 is a statistical measure that assesses the strength and direction of the linear relationship between two continuous variables. It ranges from -1 to 1, where:

* 𝑟 =1 indicates a perfect positive linear relationship.
* 𝑟 = −1 indicates a perfect negative linear relationship.
* 𝑟 = 0 indicates no linear relationship.

# RESULTS AND DISCUSSION

## COVID-19 TABLE NUMBER OF POOR AIR QUALITY PRE AND DURING COVID-19

This discussion delves into the observed changes in the number of poor air quality days before and during the COVID-19 pandemic, highlighting the potential impacts of reduced human activity on air pollution levels, aiming to provide insights into whether social isolation indeed affects air quality.



**FIGURE 2.** Graph of Days with Poor Air Quality over 12 Quarters

**FIGURE 2** illustrates the number of days with poor air quality from 2018 to 2020. The graph reveals a notable decrease in the number of poor air quality days in 2020, coinciding with COVID-19 pandemic and the implementation of social isolation in Indonesia. Based on graph there is a 15.85% decrease of poor air quality between 2018 to 2019. And 55.07% decrease of poor air quality between 2019 and 2020. A decrease of up to 50% in poor air quality would generally indicate a significant improvement in air quality. This improvement suggests that measures taken, such as reduced emissions from human activities like transportation and industry during periods of social isolation or lockdowns, have effectively reduced pollutants in the air, leading to cleaner and healthier air conditions.

In addition to prove this correlation, the author calculated the percentage change in the number of vehicles operating from 2018 to 2020, providing a comprehensive analysis of the pandemic's impact on transportation patterns.

*A graph showing the number of cars mobility

Description automatically generated*

**FIGURE 3**. Graph of Cars Mobility over 12 Quarters

**FIGURE 3** shows a significant decrease in the number of cars mobility from 2019 to 2020. Based on the calculations conducted, vehicle mobility decreased by 38.02% during this period. As depicted in the attached graph, air quality improved in 2020. This suggests a potential correlation between poor air quality and vehicle mobility. To explore this hypothesis further, the author will analyse the correlation using the Pearson method.

## PEARSON CORRELATION ANALYSIS BETWEEN NUMBER OF CARS AND AIR QUALITY PEARSON

The Pearson Correlation Model was used to measure the strength and direction of the linear relationship between the Number of Cars Mobility and Air Quality. This analysis aims to assess whether the Number of Cars can significantly increase Poor Air Quality.

Scipy.stats library were used to calculate the Pearson coefficients correlations value of the variables. Table I will present in **TABLE 3** results:

**TABLE 3**. Pearson Correlation Analysis between Number of Cars and Air Quality

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Pairs | Coefficient Correlation | Correlation Level | p-value |
| Cars Mobility and Poor Air Quality | 0.401 | Moderate | 0.195 |

The table shows that there’s a positive relationship between Number of Cars and quantity of Poor Air quality with a correlation coefficient of 0.40. There is a decrease of 55.07% of air quality in 2020. Due to the moderate correlation between Number of Cars and Air Quality, it implies that there is another factor that may have played a role to the improvement in air quality in 2020. And p-value of 0.19 is higher than common significance levels such as 0.05 or 0.1. This suggests that the evidence is not strong enough to reject the null hypothesis. This could happen because of the correlation levels. If relationship between variables is weak, the p-value may be higher [13].

Although the correlation is moderate, it does supports our hypothesis This shows that limit traffic flows seem to be effective in improving air quality [14], necessitating consideration of additional influencing factors such as rainfall intensity. Based on previous studies, rainfall intensity is considered a key factor influencing air quality improvement.

## PEARSON CORRELATION ANALYSIS BETWEEN RAINFALL INTENSITY AND AIR QUALITY

Previous studies have suggested that rainfall intensity can improve air quality by washing out pollutants from the atmosphere, a process known as wet deposition. The direct effect of rainfall is mostly the washing effect to decrease the pollution concentration[15]. Conversely, some studies indicate that heavy rainfall can contribute to soil erosion and the subsequent release of particulates into the air, potentially deteriorating air quality [16].

This mixed evidence requires a comprehensive analysis to understand the multifaceted relationship between rain and air quality. By employing the Pearson correlation method, this study aims to elucidate the relationship between rainfall and air quality, providing empirical evidence to support or refute existing theories.

Due to data loss in the rainfall intensity data specifically for Quarter 3 of 2019, the author performed reprocessing. The author addressed this issue by aggregating data from the same quarters of other years and averaging the rainfall intensity data from Q3 2018 and Q3 2020 to estimate and restore the missing data of Q3 2019 and ensure comprehensive data coverage and accuracy. **FIGURE 4** shows the graph of rainfall intensity, which we will later compare with data on good air quality and poor air quality to determine whether rainfall intensity improves or degrades air quality.

*A graph with numbers and a line

Description automatically generated*

**FIGURE 4**. Graph of Rainfall Intensity over 12 Quarters

As we can see from the **FIGURE 4**, it was discovered that rainfall intensity in 2020 was significantly higher compared to the previous eight quarters. Specifically, rainfall intensity in 2020 increased by 61% compared to the four quarters of 2019.

After applying the Pearson correlation method using the Scipy.stats library, we have compiled the results in the **TABLE 4** below:

**TABLE 4**. Pearson Correlation Analysis between Rainfall Intensity and Air Quality Results

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Pairs | Coefficient Correlation | Correlation Level | p-value |
| Rainfall Intensity & Good Air Quality | 0.756 | Strong | 0.005 |
| Rainfall Intensity & Poor Air Quality | -0.478 | Moderate | 0.121 |

After comparing rainfall quality with the two air quality variables, we obtained that there is a strong positive relationship between rainfall intensity and good air quality with significant value. While a moderate negative relationship was found between rainfall intensity and poor air quality. This result has indicated that higher rainfall intensity is linked to better air quality and lower rainfall intensity is linked to poorer air quality.

The 61% increase in rainfall intensity combined with a 38% decrease in vehicle mobility in 2020 contributed to improved air quality that year.

## PEARSON CORRELATION BETWEEN O3 AND AIR QUALITY

In this section, we present the results of the highest substances of air pollutant from the analysis of poor air quality that categorized as "UNHEALTHY" and "VERY UNHEALTHY" for each quarter from quarter 1 2018 until quarter 4 2020. The following graph shows the highest substances of air pollutant along with the average of occurrences for both "UNHEALTHY" and "VERY UNHEALTHY" air quality category.

**

**FIGURE 5**. Graph of Dominant Substances over 12 Quarters

The results in **FIGURE 5** shows that over a twelve consecutive quarters, Ozone (O3) is continuously identified as the highest pollutant substances in poor air quality. Ozone is known to pose significant risks to respiratory health, particularly through short-term exposure which can lead to acute respiratory infections (ARI) [17]. Additionally, we used Pearson correlation analysis (**TABLE 5**) to confirm the relationship between Ozone (O3) levels and worsening air quality.

**TABLE 5**. Pearson Correlation Analysis between O3 and Air Quality Results

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Pairs | Coefficient Correlation | Correlation Level | p-value |
| O3 & Poor Air Quality | 0.983 | Strong | 0.005 |

The presence of ozone in poor air quality conditions shows that ozone is a major pollutant that impacts public health. Moreover the vehicles mobility in DKI Jakarta is pretty high. The emissions consist of Ozone (O3), Oxides of sulphur (SOx), Oxides of nitrogen (NOx), Carbon dioxide (CO2), Carbon monoxide (CO), hydrocarbons and particulate matter [18]. This correlation underscores the critical importance of monitoring ozone levels in air quality management strategies. Continuous inhalation of ozone can threat the public health.

## PEARSON CORRELATION ANALYSIS BETWEEN ARI CASES AND AIR QUALITY

Short-term exposure to ambient ozone (O3) poses a risk for acute respiratory effects [19]. Based on the previous findings, there is a significant presence of ozone (O3) when the air quality is poor. Consequently, the more days with poor air quality, the higher the concentration of ozone in the air, which can lead to increased levels of acute respiratory infections (ARI). Ozone exposure can exacerbate respiratory conditions such as asthma and increase the respiratory infections, especially in vulnerable populations such as children, the elderly, and individuals with pre-existing respiratory diseases.

Understanding ozone's health impacts informs environmental policies and regulations aimed at reducing ozone levels, which is crucial for protecting public health and enhancing air quality standards.This part will explore the correlation between poor air quality and acute respiratory infection (ARI) cases using the Pearson correlation method.

**TABLE 6**. Pearson Correlation Analysis between ARI Case and Air Quality Results

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Pairs | Coefficient Correlation | Correlation Level | p-value |
| ARI Case & Poor Air Quality | 0.365 | Weak | 0.36 |

Based on the analysis we have conducted in **TABLE 6**, the coefficient correlation between ARI Case and Poor Air Quality is 0.36. This means that there is a low positive correlation between ARI cases and poor air quality. From this result, we can tell that if poor air quality increases, there is also a tendency for ARI cases to increase. Although the correlation is not too strong, it is still important to consider that air quality can affect humans respiratory health.

Understanding this relationship is crucial for public health initiatives and policymaking. Even though the correlation is not too strong, it still highlights the need to consider air quality as a factor in health strategies aimed at reducing respiratory illnesses. Further research is necessary to delve deeper into the underlying mechanisms and factors contributing to these correlations. This ongoing investigation can provide insights into how environmental regulations and interventions can be optimized to mitigate the health impacts of poor air quality effectively.

# CONCLUSIONS

This research offers a comprehensive analysis of poor air quality trends before and during the COVID-19 pandemic. The findings from 2018 to 2020 highlight a noticeable improvement in air quality linked to reduced vehicle mobility during periods of social isolation of COVID-19. Employing the Pearson correlation method, it revealed a moderate positive correlation coefficient of 0.401 between vehicle mobility and air quality. This correlation underscores the impact of reduced human activity, particularly vehicular traffic, on enhancing air quality conditions. Due to moderate correlation, we include Rainfall intensity as additional factor based on previous studies that state Rainfall intensity has long been debated for its dual impact on air quality. The findings indicated a strong positive correlation of 0.756 between higher rainfall intensity and improved air quality, which indicates that higher rainfall intensity leads to improve air quality. This research also find that ozone (O3) consistently emerges as the primary contributor to poor air quality over three consecutive years with. Based on 0.983 Pearson correlation coefficient, it does support the findings that ozone has been a consistent contributor to poor air quality. Ozone poses significant risks to respiratory health, especially through short-term exposure leading to acute respiratory infections. Additionally, the correlation analysis reveals a moderate positive relationship (correlation coefficient of 0.366) between poor air quality and ARI cases. This suggests that higher levels of poor air quality may correlate with increased occurrences of respiratory infections, highlighting the relevance of air quality considerations in public health strategies.

Due to data limitations, the analysis is limited to specific periods and variables. Also the study focus is limited on changes in air quality pre and during COVID-19, with variables selected based on pandemic conditions, such as reduced car mobility due to social isolation. Further research is needed includes other factors like industrial emissions, construction activities, and others that impact air quality with more extended period to strengthen the significance of the findings. This research can be further developed to assess the negative impact of ozone O3 on acute ARI by employing Geographic Information Systems to map ARI cases against localized O3 concentrations. This spatial analysis could reveal hotspots where high O3 levels coincide with increased ARI rates, highlighting areas for targeted interventions.

# Acknowledgments

We would like to express our gratitude to all individuals and institutions who supported this research. Special thanks to our families and friends for their continuous encouragement and understanding during the course of this study. We also appreciate the valuable feedback and guidance from our colleagues and the contributions of everyone involved in this study.

# References

1. Mekuria, G., *Air pollution: a review of its impacts on health and ecosystems, and analytical techniques for their measurement and modeling.* Journal of Environmental Informatics Letters, 2023. **10**: p. 115-131.

2. Clayton, S. and C. Manning, *Psychology and climate change: Human perceptions, impacts, and responses*. 2018: Academic Press.

3. Fath, M.A., *Pengaruh Kualitas Udara dan Kondisi Iklim terhadap Perekonomian Masyarakat (Literature Review).* Media Gizi Kesmas, 2021. **10**(2): p. 2021-329.

4. Manisalidis, I., et al., *Environmental and health impacts of air pollution: a review.* Frontiers in public health, 2020. **8**: p. 14.

5. Saurabh Sonwani, S.S. and V.M. Vandana Maurya, *Impact of air pollution on the environment and economy*, in *Air pollution: Sources, impacts and controls*. 2019, CAB International Wallingford UK. p. 113-134.

6. Xu, H., et al., *Environmental and health risks of VOCs in the longest inner–city tunnel in Xi’an, Northwest China: Implication of impact from new energy vehicles.* Environmental Pollution, 2021. **282**: p. 117057.

7. Pérez, J., et al. *A data preparation methodology in data mining applied to mortality population databases*. Springer.

8. Albashrawi, M., *Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015.* Journal of Data Science, 2016. **14**(3): p. 553-569.

9. Taylor, K.S., K.R. Mahtani, and J.K. Aronson, *Summarising good practice guidelines for data extraction for systematic reviews and meta-analysis.* BMJ Evidence-Based Medicine, 2021. **26**(3): p. 88-90.

10. Represa, N.S., et al., *Data mining paradigm in the study of air quality.* Environmental Processes, 2020. **7**(1): p. 1-21.

11. Ndako, J.A., et al., *Evaluation of the association between malaria infection and electrolyte variation in patients: use of Pearson correlation analytical technique.* Informatics in Medicine Unlocked, 2020. **21**: p. 100437.

12. Wahyuning, S., *Dasar-dasar statistik.* Penerbit Yayasan Prima Agus Teknik, 2021: p. 1-105.

13. Schober, P., C. Boer, and L.A. Schwarte, *Correlation coefficients: appropriate use and interpretation.* Anesthesia & analgesia, 2018. **126**(5): p. 1763-1768.

14. Šverko, Z., et al., *Complex Pearson correlation coefficient for EEG connectivity analysis.* Sensors, 2022. **22**(4): p. 1477.

15. Coscia, M., *Pearson correlations on complex networks.* Journal of Complex Networks, 2021. **9**(6): p. cnab036.

16. Cumming, G., *The new statistics: Why and how.* Psychological science, 2014. **25**(1): p. 7-29.

17. Rossi, R., R. Ceccato, and M. Gastaldi, *Effect of road traffic on air pollution. Experimental evidence from COVID-19 lockdown.* Sustainability, 2020. **12**(21): p. 8984.

18. Kwak, H.-Y., et al., *Identifying the correlation between rainfall, traffic flow performance and air pollution concentration in Seoul using a path analysis.* Transportation research procedia, 2017. **25**: p. 3552-3563.

19. Mohamadi, M.A. and A. Kavian, *Effects of rainfall patterns on runoff and soil erosion in field plots.* International soil and water conservation research, 2015. **3**(4): p. 273-281.