Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:9
https://doi.org/10.28919/cmbn/8982
ISSN: 2052-2541

STATISTICAL MODELING OF METEOROLOGICAL FACTORS AND PM2.5 LEVELS: IMPLICATIONS FOR CLIMATE CHANGE IN INDONESIA

RESTU ARISANTI[1,*], RESA SEPTIANI PONTOH[1], SRI WINARNI[1], VALERIE VINCENT YANG[2],
RAHMA ARIZA RIANTIKA[2], GINTA YUFENDI LAURA[2]

[1]Department of Statistics, Padjadjaran University, Indonesia

[2]Bachelor Program of Statistics Department, Padjadjaran University, Indonesia

*Correspondence author: r.arisanti@unpad.ac.id

Abstract: Increased air pollution in Indonesia has resulted in fluctuations in climate patterns. The development of this research lies in the response variable studied, namely the concentration of PM2.5 which is the main indicator and a serious concern in Indonesia because the high concentration of these particles has a negative impact on public health. The Air Pollution Standard Index (ISPU) with the PM2.5 parameter is used to measure air quality based on the amount of PM2.5 levels in the air. ISPU with the PM2.5 parameter can show two possible levels of air quality, namely dangerous or not so that air quality data seen from the ISPU level of the PM2.5 parameter is binary. Meteorological factors that are thought to affect PM2.5 concentrations are air temperature, rainfall, solar irradiation, air humidity, air pressure, and wind speed. This study uses a binary logistic regression analysis method that aims to determine the influence of meteorological factors that are thought to affect PM2.5 concentrations so that they have the potential to affect climate change in Indonesia. The results of the analysis show that Indonesia with provinces that have PM2.5 concentrations still dominate, with a dangerous category of 52.94%. The meteorological factors that have a significant

[*]Corresponding author
E-mail address: r.arisanti@unpad.ac.id
Received October 23, 2024

effect on PM2.5 concentrations are air temperature, rainfall, solar irradiation, air humidity, air pressure, and wind speed. In addition, climate change in Indonesia is also influenced by these factors. The research findings show that these characteristics have consequences for the Indonesian government's initiatives aimed at improving air quality and addressing climate change. Therefore, this is in line with the objectives of achieving the Sustainable Development Goals (SDGs), such as promoting Good Health and Wellbeing (SDG 3), creating Sustainable Cities and Communities (SDG 11), and undertaking Climate Action (SDG 13).
Keywords: $PM_{2.5}$; meteorological conditions; climate change; binary logistic regression.
2020 AMS Subject Classification: 62P12.

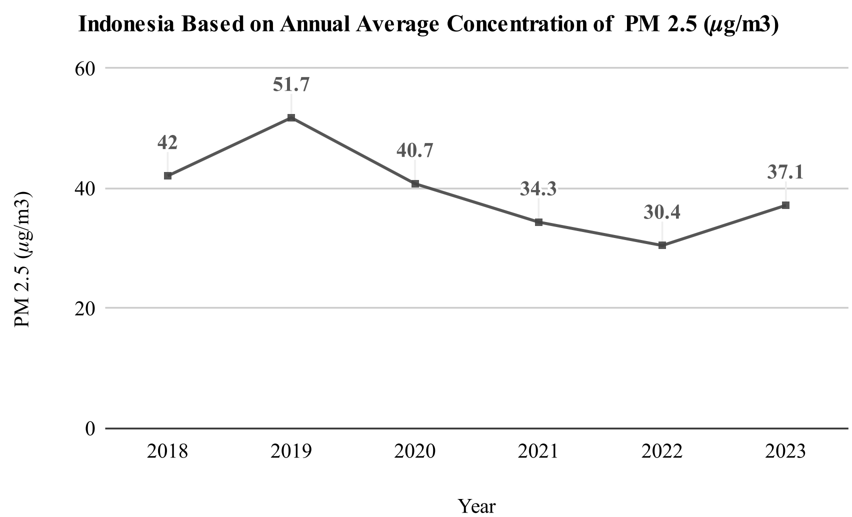
1. INTRODUCTION

Clean air is one of the components of basic needs that are important for the comfort, health, and well-being of living things on earth [1], [2] . In the era of regional autonomy and globalization, dynamic and complex developments that occur in all fields have consequences for humans and the environment [3], especially potentially affecting air quality. The movement of air quality up or down is realized from human activities and various sectors such as industry, transportation, and so on so that air pollution affects air quality [4]. The entry of harmful elements in the form of particles or gasses into the atmosphere that can cause adverse effects on living things so that changes occur in decreasing air quality is called air pollution [5]. According to IQAir 2023, Indonesia is ranked 14th for the country with the worst air quality in the world with a value of 105 which is categorized as "unhealthy for sensitive groups" in the Air Quality Index (AQI). Air pollution has an impact on respiratory disorders and non-carcinogenic risks in society [6]. Thus, the serious adverse impacts of air pollution in Indonesia are at the center of public attention and scientific studies.

The concentration of $PM_{2.5}$ (particulate matter with an aerodynamic diameter of less than 2.5 μm) is one of the most harmful components of several types of air pollutants [7] that can cause serious health problems and even death if inhaled [8]. Based on C40 Knowledge, the World Health Organization (WHO) guidelines state that the annual average concentration of $PM_{2.5}$ should not exceed 5 $\mu g/m^3$, while the 24-hour average exposure should not exceed 15 $\mu g/m^3$ and not exceed 3-4 days per year. In Figure 1, the annual average concentration of $PM_{2.5}$ in Indonesia ranges from

6 to 10 times compared to the limitations of the WHO guidelines.

In line with the United Nations Sustainable Development Goals (SDGs), particularly SDG 3 (Good Health and Well-Being) and SDG 11 (Sustainable Cities and Communities), addressing $PM_{2.5}$ pollution is crucial for promoting healthier living conditions and sustainable urban development. Moreover, reducing air pollution is integral to achieving SDG 13 (Climate Action) as it contributes to mitigating climate change impacts by reducing greenhouse gas emissions and improving air quality.



Source: IQAir

Figure 1. Indonesia by Annual Average $PM_{2.5}$ Concentration (µg/m^3)

In summary, high $PM_{2.5}$ concentrations can have significant negative impacts on human health, climate change, local ecosystems and economic development [9]. The relationship between $PM_{2.5}$ concentrations and climate change are a highly complex and multifaceted one that is influenced by various factors, including meteorological conditions, emission sources, and seasonal patterns. Among the influencing factors, meteorological conditions are among the most important [10]. The adverse impacts of elevated $PM_{2.5}$ levels are not only limited to air quality issues as they have the potential to affect regional and global climate patterns [11] . This study uniquely examines the concentration of $PM_{2.5}$ as the response variable, distinguishing it from previous research which focused on other pollutants like ozone (O_3). By doing so, it provides new insights into the interplay between air pollution and climate change specific to Indonesia, highlighting the unique meteorological and socio-economic factors influencing $PM_{2.5}$ levels. The study of the relationship between meteorological conditions and $PM_{2.5}$ concentrations can help us to have a better understanding of the $PM_{2.5}$ air quality problem, and can contribute to the implementation of more effective measures to reduce $PM_{2.5}$ concentrations. In Figure 2, the distribution of $PM_{2.5}$ concentrations dominates in Indonesia's metropolitan areas compared to other areas which are occupied by a lot of people and industrial activities, office buildings, and so on.

Air quality based on $PM_{2.5}$ concentration levels is measured every day for 24 hours continuously through climatology stations spread across all provinces in Indonesia and can be seen based on the Air Quality Standard Index (ISPU) with $PM_{2.5}$ parameters. ISPU calculations are made based on the upper limit ISPU, lower limit ISPU, upper limit ambient, lower limit ambient, and ambient concentration of measurement results. Based on IQAir at the US AQI Level, the ISPU level of the $PM_{2.5}$ parameter will be 1 if the $PM_{2.5}$ concentration in the air is at a dangerous level for an individual living being where the ISPU value is more than 12.1 and is 0 otherwise. Thus, air quality based on the ISPU level of the $PM_{2.5}$ parameter is binary data. Meteorological conditions such as air temperature, rainfall, sunlight, humidity, air pressure, and wind speed will be used as predictor variables that can affect the air quality considered from the ISPU level of the $PM_{2.5}$ parameter [12].

Figure 2. Distribution of $PM_{2.5}$ Concentrations (g/m^3) in Indonesia in 2022

Binary logistic regression is a special type of regression that belongs to the class of non-linear regression. It is used specifically when the response variable has two or more categories (success or failure) [13]. In this study, we conducted a binary logistic regression analysis in 34 provinces in Indonesia based on 1-year observation records in 2022.

By investigating the specific meteorological factors influencing $PM_{2.5}$ concentrations and their implications for air quality, this study advances our understanding of how local climatic conditions interact with pollution dynamics, providing critical information for policy interventions aimed at achieving cleaner air and aligning with SDG targets. This focus on $PM_{2.5}$, as opposed to other pollutants studied in prior research, underscores the novel approach and scope of this study, offering valuable insights for both environmental policy and public health initiatives in Indonesia. Factors in meteorological conditions will be analyzed using binary logistic regression on the response variable, namely, the ISPU level of $PM_{2.5}$ parameters. The goal is to determine whether there are significant differences that affect the research in predicting the relationship between meteorological conditions as predictor variables (x) and $PM_{2.5}$ concentrations as response variables (y) to get an explanation of climate change in Indonesia. The benefits of the conclusions of this study can be used to improve understanding of the mechanism of meteorological conditions and the accuracy of forecasts of ISPU $PM_{2.5}$ parameters related to climate change and can be a reference in making environmental and air management policies in Indonesia.

2. PRELIMINARIES

Nationally, air pollution is a major risk factor for non-communicable diseases in Indonesia. It is the largest environmental cause of disease burden in the country [14]. For instance, the total annual cost of health impacts from air pollution in the capital city of Indonesia which named Jakarta is approximately $2.9 billion [15]. $PM_{2.5}$ exposure is estimated to cause over 7,000 adverse health outcomes in children, over 10,000 premature deaths, and over 5,000 hospitalizations annually in Jakarta alone. [15].

Increasing temperatures will likely enhance photochemical production of secondary $PM_{2.5}$[16]

[14]. Increasing temperatures will likely enhance photochemical production of secondary $PM_{2.5}$ [16] [14]. Changes in precipitation patterns may alter wet deposition of $PM_{2.5}$ and precursor gases [16]. Shifts in wind patterns could impact the transport and dispersion of air pollutants [17]. More frequent extreme weather events like droughts may increase the frequency and severity of $PM_{2.5}$ episodes [16].

Meteorology has a significant influence on $PM_{2.5}$ levels, with temperature, wind, and precipitation being key factors. High $PM_{2.5}$ exposure leads to major health and economic burdens. Climate change has the potential to exacerbate air pollution issues in the region through various mechanisms. Integrated analysis of air quality and meteorological data is needed to better understand these interactions and develop effective mitigation strategies. [16] [18].

Previous research in China shows the interactions between meteorological factors, such as temperature, wind speed, wind direction, humidity, precipitation, radiation, atmospheric pressure, and planetary boundary layer height, with $PM_{2.5}$ concentrations through various mechanisms, including dispersion, growth, chemical production, photolysis, and deposition of $PM_{2.5}$. This study provides comprehensive insights into the variations in meteorological influences on $PM_{2.5}$ concentrations across different regions and highlights the feedback effects of $PM_{2.5}$ concentrations on these meteorological factors. These findings suggest that strong bidirectional interactions between $PM_{2.5}$ and meteorological factors are a key driver in exacerbating $PM_{2.5}$ pollution [19].

Multicollinearity Test

Multicollinearity refers to the presence of a significant linear relationship between two or more predictor variables in regression analysis [20]. One method to check the multicollinearity assumption is by the Variance Inflation Factor (VIF) value on each predictor variable. The VIF formula is as follows.

$$VIF = \frac{1}{(1-{R_j}^2)} \tag{1}$$

${R_j}^2$ represents the regression determination coefficient between the jth predictor variable and the remaining predictor variables (k-1). If the VIF value > 10, it indicates the presence of multicollinearity. Thus, it can be concluded that the assumption about the absence of multicollinearity is met if the VIF value is < 10.

Binary Logistic Regression

Binary logistic regression is one of the specific models of generalized linear model (GLM). GLM does not require classical regression assumptions [21]. In data analysis, binary logistic regression can be used to generate a model in determining the relationship between response variables (y) that are binary (dichotomous) with several predictor variables (x) that are categorical or continuous [22], [23]. If the response variable (y) in logistic regression analysis consists of two categories, such as $Y = 1$ indicates the result obtained is "success" and $Y = 0$ indicating the result obtained "failure", then the logistic regression uses binary logistic regression [24]. The response variable (y) in binary logistic regression follows a Binomial distribution or Bernoulli distribution [25], then the probability function y [26] is as follows.

$$f(Y = y) = \pi(x)^y (1 - \pi(x))^{1-y}, y = 0,1 \tag{2}$$

"Success" and "failure" are two criteria generated by the response variable (y) denoted by $Y = 1$ represents the probability of success then $P(Y = 1) = \pi$ and $Y = 0$ representing the probability of no success/failure then $P(Y = 0) = 1 - \pi$. The binary logistic regression model used [27], is as follows.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_6 x_6}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_6 x_6}} \tag{3}$$

$\pi(x)$ is the probability of a successful event with a probability value $0 \leq \pi(x) \leq 1$ where β_0 is an intercept in the form of a constant number, $\beta_1, \ldots, \beta_6$ are logistic regression parameters and $x_{1,\ldots,} x_6$ which is the value of the independent variable. Model with k which is the number of independent variables. In logistic regression, the dependent variable (Y) and independent variables (X) are not modeled directly. By transformation, the dependent variable is converted into a logit variable, which is the *natural log of* the *odds ratio* so that equation (2) is transformed using logit as follows.

$$g(x) = log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_6 x_6 \tag{4}$$

To perform parameter estimation β with $\beta = (\beta_1, \beta_2, \ldots, \beta_6)$, the Maximum Likelihood Estimation (MLE) method is used, which will be discussed further.

Maximum Likelihood Estimation (MLE)

Parameter estimation in binary logistic regression models can use the Maximum Likelihood Estimation (MLE) method [28]. The weighted least squares method with several iteration processes is referred to as the Maximum Likelihood Estimation (MLE) method [29]. The method estimates the parameter β by maximizing the likelihood function and requires that the data must follow a certain distribution [27]. If the dependent variable (Y) is changed to 0 or 1 then the statement is a logistic regression equation with conditional probability $Y = 1$ given x is denoted as $P(Y = 1|x)$, then the quantity $1 - \pi(x)$ by giving the conditional probability $Y = 0$ given x is denoted as $P(Y = 0|x)$. Furthermore, for (x_i, y_i), where $y_i = 1$ the involvement for the likelihood function is $\pi(x_i)$ and for $y_i = 0$ the involvement for the likelihood function is $1 - \pi(x_i)$, where the quantity $\pi(x_i)$ produces the value $\pi(x)$ at x_i. Then the maximum likelihood function is as follows.

$$L(\beta) = \prod_{i=1}^{n} f(x_i; \beta) \tag{5}$$

$$L(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{6}$$

With y_i is the observation of the *i-th* explanatory variable and $\pi(x_i)$ is the probability for the *i-the explanatory variable*. In equation (6), the logit approach is used to estimate the parameters so that the *log-likelihood* function is as follows.

$$l(\beta) = ln[L(\beta)] = \sum_{i=1}^{n}\{y_i\, ln[\pi(x_i)] + (1 - y_i)\, ln[1 - \pi(x_i)]\} \tag{7}$$

$$l(\beta) = \sum_{j=0}^{k}\left[\sum_{i=1}^{n} y_i\, x_{ij}\right]\beta_j - \sum_{i=1}^{n} ln\left[1 + exp(\sum_{j=0}^{k} \beta_j x_{ij})\right] \tag{8}$$

The expected value of the parameter β is obtained by making the first derivative of $l(\beta)$ against β, the result obtained is equal to zero. From this result, the estimator is obtained $\pi(x_i)$ where $g(x)$ is the logit estimator as a linear function of the explanatory variables [27]. Parameter values β is calculated by finding the first partial derivative of each parameter followed to maximize the likelihood function as follows.

$$l\frac{\partial l(\beta)}{\partial(\beta_1)} = 0; l\frac{\partial l(\beta)}{\partial(\beta_2)} = 0; l\frac{\partial l(\beta)}{\partial(\beta_3)} = 0; l\frac{\partial l(\beta)}{\partial(\beta_4)} = 0; l\frac{\partial l(\beta)}{\partial(\beta_5)} = 0; l\frac{\partial l(\beta)}{\partial(\beta_6)} = 0 \tag{9}$$

Often, the first derivation of equation (8) does not produce clear results, so equation (6) is optimized by the Newton Raphson numerical method.

Parameter Significance Test

The significance test of loading parameters is carried out simultaneously (*overall*) and partially will be applied to all parameter coefficients of independent variables on the response/dependent variables.

To evaluate and test the effect of the coefficient β coefficients in the model, a simultaneous test is carried out using the following hypothesis.

$H_0: \beta_1 = \beta_2 = \cdots = \beta_6 = 0$; the independent variables simultaneously do not have a significant effect on the response variable.

H_1: there is at least one $\beta_j \neq 0$ with $j = 1,2,\ldots,6$; independent variables simultaneously have a significant effect on the response variable.

The simultaneous test statistic used for the model is [30] as follows.

$$G^2 = -2ln\left[\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right] \tag{10}$$

Simultaneous testing is done using the Chi-square test with the formula equation (10) following the Chi-square distribution G^2 with degrees of freedom $k$where the test criterion is reject H_0 if $G > X^2_{(k;\alpha)}$ or $p - value < \alpha$, accept in other cases.

Partial testing is done by comparing standard errors to determine the effect and influence of the independent variable parameters individually. The hypothesis for the partial test is as follows.

$H_0: \beta_1 = \beta_2 = \cdots = \beta_6 = 0$; the independent variable j does not have a significant effect on the response variable.

H_1: $\beta_j \neq 0$ with $j = 1,2,\ldots,6$; the independent variable$j$ has a significant effect on the response variable.

The partial test statistics used for the model are as follows.

$$Z = \frac{\widehat{\beta}_j}{SE(\widehat{\beta}_j)} \tag{11}$$

Partial testing is done using the Wald test with the formula equation (11) which can be approximated by the Normal distribution where $\hat{\beta}_j$ is the estimator β_j while $SE(\hat{\beta}_j)$ is the standard error. The test criteria for the Wald test are reject H_0 if $|Z| > Z_{(1;\alpha)/2}$ or $p - value <$

α, accept in other cases.

Model Fit Test

The model fit test is used to evaluate whether the model fits the data, the observed values obtained are the same or close to those expected in the model. The tool used to test the fit in binary logistic regression is the *deviance* test.

The deviance statistic is formulated as follows.

$$D = -2\sum_{i=1}^{n}\left\{y_i\ ln\left(\frac{n_1\hat{\pi}_1}{y_i}\right) + (1-y_i)\ ln\ \left(\frac{n_1-n_1\hat{\pi}_1}{n_1-y_i}\right)\right\} \tag{12}$$

The test criteria for the deviance test is reject H_0 if $D > X_{\alpha,df}$ or $p-value < \alpha$, accept in other cases.

Cross-Sectional Data

Cross-sectional data is data obtained from cross-sectional research, where the phenomenon is observed at one specific point in time. In this type of research, the researcher only examines the phenomenon at one particular moment without involving changes over time. In exploratory, descriptive, or explanatory research, cross-sectional research can explore the relationship between observed variables in a given population, test models or hypotheses, and compare sample groups at the same point in time. However, cross-sectional research is unable to capture the dynamics of change over time in the observed population, or the variables that change over time that affect it. Thus, cross-sectional data is data collected at the same point in time from different regions or subjects [31][32].

Data

The data used in this research study uses secondary data on air pollution against climate change in 34 provinces spread across Indonesia in terms of the Air Pollution Standard Index (ISPU) with $PM_{2.5}$ parameters obtained from the official publication of the Central Statistics Agency (BPS) in 2022. The data set consists of 34 observations which is the average or the number of events (meteorological condition factors) that occur during 1 year, namely 2022. This study will discuss the factors of meteorological conditions that affect the level of ISPU with $PM_{2.5}$ parameters in Indonesia in 2022 using binary logistic regression analysis. This research was conducted with the help of R *software* and Microsoft Excel.

In this study, the ISPU with the $PM_{2.5}$ parameter will act as an independent variable/response (y), where the ISPU level with a dangerous category is denoted by a value of 1 and a value of 0 if the ISPU level is categorized as safe. The independent variables in this study are air temperature (x_1) with units of Celsius, rainfall (x_2) with units of mm/year, sunlight (x_3) with units of hours, humidity (x_4) with units of percent, air pressure (x_5) with units of millibars, and wind speed (x_6) with units of knots.

Data Processing

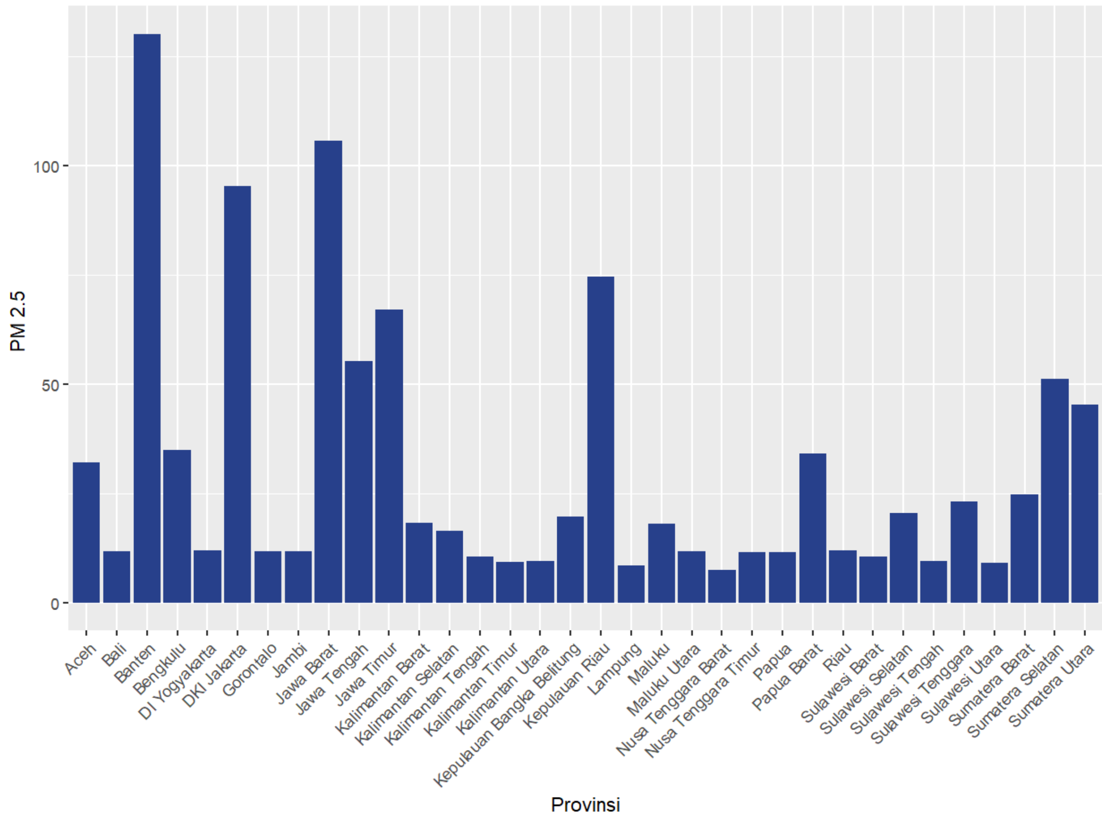
The steps to build a binary logistic regression model on *cross-sectional* data are as follows.

1) Data exploration on *cross-sectional data* to determine the characteristics of the data in determining the initial structure of the model.
2) Perform an assumption test, namely the multicollinearity test as in equation (1)
3) Estimating logistic regression parameters using *Maximum Likelihood Estimation* (MLE)
4) Testing the significance of parameters simultaneously with the t-test$G$ as in equation (10)
5) Partial parameter significance testing with the Wald test as in equation (11)
6) Formation of binary logistic regression model
7) Testing the fit of the model using the *deviance* test as in equation (12)
8) Making conclusions

3. MAIN RESULTS

Descriptive statistics

Before conducting binary logistic regression analysis on the data, we will start by conducting descriptive statistical analysis to determine the characteristics of provinces in Indonesia based on the ISPU level of airborne $PM_{2.5}$ parameters.



Source: IQAir, 2022

Figure 3. $PM_{2.5}$ Concentration ($\mu g/m^3$) in 34 Provinces in Indonesia in 2022

In Figure 3, shows that in 2022, Banten Province recorded the highest $PM_{2.5}$ concentration of 130 $\mu g/m^3$ US AQI Index categorized as *unhealthy*. West Java Province ranked second with a $PM_{2.5}$ concentration of 105.6 $\mu g/m^3$ in the US AQI Index categorized as *unhealthy*, and DKI Jakarta Province ranked third with a $PM_{2.5}$ concentration of 95.3 $\mu g/m^3$ in the US AQI Index categorized as *unhealthy*. For the lowest $PM_{2.5}$ concentration of 7.4 $\mu g/m^3$ is West Nusa Tenggara Province in the US AQI Index categorized as *"Good"*.

Table 1. Descriptive Analysis

Variable	Average	Minimum	Maximum	Variety
$PM_{2.5}$	30.4	7.9	130	956.1445

Source: R program's results

Table 1, shows that the average concentration of $PM_{2.5}$ in Indonesia in 2022 reached 30.4 μg/m^3 indicating that in general, the provinces in Indonesia can be grouped into the US AQI Index *moderate* category. The variance value of 956.1445 indicates that the $PM_{2.5}$ (μg/m^3) data in 34 provinces in Indonesia in 2022 is widely dispersed from the average value. In addition, the range value of 122.1 shows that there is a large difference between the provinces in Indonesia with the highest and lowest $PM_{2.5}$ concentrations.

Table 2. Descriptive Analysis

Variable	Category	Percentage	Total
$PM_{2.5}$	0 = "Safe"	47.06	100
	1= "Dangerous"	52.94	

In Table 2, it shows that Indonesia is dominated by provinces that have $PM_{2.5}$ concentrations included in the US AQI Index categorized as dangerous with a percentage of 52.94%.

Multicollinearity Test

The results of multicollinearity testing with formula (2.1) on all independent variables, namely meteorological condition factors, obtained VIF values of 4.479426, 1.030812 1.271802, 2.393491, 3.509914, and 1.350100 respectively for x_1, x_2, x_3, x_4, x_5 and x_6 Where these values are smaller than 10. Based on the VIF value < 10, it can be concluded that accept H_0. This means that there is no multicollinearity in the independent variables. This means that the variables of air temperature, rainfall, sunlight, humidity, air pressure, and wind speed are not correlated with each other.

Parameter Estimation

Cross-sectional data analysis of the ISPU level based on $PM_{2.5}$ parameters is carried out to form a binary logistic regression model specification that will use the Maximum Likelihood Estimation (MLE) method to estimate parameters. β. The parameter estimation results obtained will be presented as follows.

Table 3. Binary Logistic Regression Parameter Estimation Results

Coefficient	Parameters	Estimation	Std. Error	z value	Pr(>|z|)	Exp (Coef.)
(Intercept)	$\hat{\beta}_0$	119.1	59.4	2.005	0.0450*	5.27e+51
x_1	$\hat{\beta}_1$	4.297	1.706	2.519	0.0118*	73.44691
x_2	$\hat{\beta}_2$	0.00149	0.000713	2.090	0.0366*	1.00149
x_3	$\hat{\beta}_3$	-2.194	0.9999	-2.195	0.0282*	0.111427
x_4	$\hat{\beta}_4$	0.4686	0.2259	2.075	0.0380*	1.597833
x_5	$\hat{\beta}_5$	-0.2736	0.1082	-2.527	0.0115*	0.760661
x_6	$\hat{\beta}_6$	1.386	0.5982	2.317	0.0205*	3.998649

In Table 3, all parameters have a value of $p - value$ which is smaller than the specified significance value, namely $\alpha = 5\%$ so it can be concluded that all parameters are significant to the model.

To determine the magnitude of the influence of each significant independent variable in the study, it can be known through the interpretation of the coefficient or what is known as the *odds ratio*. In Table 3, the interpretation of the coefficient obtained states that high air temperature has a 73.45 times greater risk of making the ISPU level of the $PM_{2.5}$ parameter at a dangerous level compared to normal air temperature. Furthermore, high wind speed provides a 4 times greater risk of making the ISPU level of the $PM_{2.5}$ parameter at a dangerous level compared to the wind speed under normal conditions. In addition, high humidity provides a 1.6 times greater risk of making the ISPU level of the $PM_{2.5}$ parameter at a hazardous level compared to humidity under normal conditions. Moreover, low sunlight provides a 0.11 times greater risk of making the ISPU level of the $PM_{2.5}$ parameter at a hazardous level compared to sunlight under normal conditions. Similarly, low air pressure provides a 0.76 times greater risk of making the ISPU level of the $PM_{2.5}$ parameter at a dangerous level compared to air pressure under normal conditions. There is no possibility that the ISPU level of the $PM_{2.5}$ parameter will change as the values of these variables increase if the coefficient (*odd ratio*) *of these* variables is equal to 1.

Simultaneous Test (Overall)

Simultaneous test results with a statistical test value G obtained equal to 19.192 where the value is greater than the value of $X^2_{(6;0.05)}$. Then, the value $p-value$ obtained at 0.003852 where the value is smaller than the specified significance value, namely $\alpha = 5\%$. Based on the value of $G > X^2_{(6;0.05)}$ and $p-value < \alpha$ then it can be concluded that rejecting H_0. This shows that overall and simultaneously the factors of meteorological conditions in the form of air temperature, rainfall, sunlight, humidity, air pressure, and wind speed in the study affect the response variable, namely the ISPU level of $PM_{2.5}$ parameters significantly. Since the factors of meteorological conditions affect significantly, it will also potentially affect climate patterns in Indonesia.

Partial Test

Based on the results of partial testing of binary logistic regression parameter estimates contained in Table 3 with a significance level of 5%, it is found that all factors of meteorological conditions and constants that affect the ISPU level of the $PM_{2.5}$ parameter significantly include the air temperature variable, with a statistical test value of 0.05. (x_1) with a statistical test value Z of 2.519, rainfall variable (x_2) with a value Z of 2.090, sunlight variable (x_3) with a value Z of -2.195, humidity variable (x_4) with a value Z of 2.075, air pressure variable (x_5) with a value Z of -2.527, and wind speed variable (x_6) with a value Z of 2.317. From these values it is known that $|Z| > Z_{(1;a)/2}$ with $\alpha = 5\%$. Moreover, the value of $p-value$ of each variable x_1, x_2, x_3, x_4, x_5 and x_6 are 0.0118, 0.0366, 0.0282, 0.0380, 0.0115, and 0.0205 respectively which are less than $\alpha = 0.05$. Since $|Z| > Z_{(1;a)/2}$ or $p-value < \alpha$ it can be concluded that rejecting H_0. This indicates that air temperature, rainfall, sunlight, humidity, air pressure, and wind speed partially affect the ISPU level of the $PM_{2.5}$ parameter.

Binary Logistic Regression Model

The binary logistic regression model obtained from the parameter estimation results is as follows.

$$\pi(x) = \frac{exp\ g(x)}{1+exp\ g(x)} \tag{13}$$

with the logit model resulted in the following model.

$$g(x) = (119.1) + (4.297)x_1 + (0.00149)x_2 - (2.194)x_3 + (0.4686)x_4 - (0.2736)x_5 + (1.386)x_6$$

Based on the logit model above, it shows that the effect of each predictor variable (x) is different. The higher the air temperature in Indonesia, the higher the ISPU level with the $PM_{2.5}$ parameter. Then, if the frequency of rainfall in Indonesia increases, the ISPU level will also increase with the $PM_{2.5}$ parameter. In addition, the lower the intensity of sunlight will make the ISPU level with the $PM_{2.5}$ parameter increase. Moreover, the higher the air humidity in Indonesia, the higher the ISPU level with the $PM_{2.5}$ parameter. In addition, the higher the air pressure in Indonesia, the lower the ISPU level with the $PM_{2.5}$ parameter. Also, high wind speed will make the ISPU level with $PM_{2.5}$ parameter increase. If the ISPU level with a high $PM_{2.5}$ parameter will also potentially affect climate patterns in Indonesia.

Model Fit Test

The logistic regression model fit test is used to evaluate how well the model fits the observed data. To see the results of whether the data in the logistic regression model is suitable, it can be seen by using the *deviance* value.

Table 4. Binary Logistic Regression Deviance Test Result

Deviance	df	p-value
27.82469	27	0.4200544

Based on Table 4, shows the results obtained from the *deviance* value using Chi-Square of 27.82469 with a *df* of 27 and the value of $P\text{-}value$ (0.4200544) > α (0.05) meaning that it fails to reject $H0$. With a confidence level of 95% the existing data fails to reject $H0$ which means the model fits the data. The binary logistic regression model used has a good fit with the observed data, so the model can be considered adequate to explain the relationship between predictor variables and response variables.

4. CONCLUSION

Based on the analysis and modelling results, meteorological factors are shown to influence the level of ISPU with the parameter PM2.5. Meteorological factors are also proven to influence climate change in Indonesia. This modelling analysis uses binary logistic regression with parameter estimation, namely the Maximum Likelihood Estimation (MLE) method. The results of the analysis show that the independent variables, which include meteorological factors such as air temperature, rainfall, sunshine, humidity, air pressure, and wind speed, significantly influence the level of ISPU with the parameter PM2.5. The effect is significant at the 5% significance level. This finding confirms that meteorological conditions play an important role in air quality variability, especially in relation to PM2.5 concentrations.

Furthermore, the significant influence of meteorological factors on PM2.5 ISPU levels indicates the potential impact of climate change in Indonesia. The complex relationship between PM2.5 concentrations and climate change suggest a mutually influential relationship. Local climate conditions can influence the distribution and concentration of air pollutants, including PM2.5. The results of this analysis have implications for government policies to improve air quality and mitigate climate change in Indonesia.

Therefore, this study makes an important contribution to the understanding of the interaction between meteorological conditions and air quality, and its relevance to climate change in Indonesia. The results of this analysis can be used as a basis for formulating environmental policies and managing air quality more effectively. Furthermore, in an effort to achieve sustainable development goals (SDGs), such as Good Health and Wellbeing (SDG 3), Sustainable Cities and Communities (SDG 11), and Climate Action (SDG 13).

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Directorate of Research and Community Engagement of Padjadjaran University for their valuable support during the writing of this paper and for providing financial support for publication in this journal.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] D. Kartikasari, Nalisis Faktor-Faktor Yang Mempengaruhi Level Polusi Udara Dengan Metode Regresi Logistik Biner, MATHunesa J. Ilm. Mat. 8 (2020), 55–59. htps://doi.org/10.26740/mathunesa.v8n1.p55-59.

[2] J. Luo, P. Du, A. Samat, J. Xia, M. Che, Z. Xue, Spatiotemporal Pattern of PM2.5 Concentrations in Mainland China and Analysis of Its Influencing Factors Using Geographically Weighted Regression, Sci. Rep.7 (2017), 40607. https://doi.org/10.1038/srep40607.

[3] D.R. Arini, C. Purnawan, E.S. Rahayu, N.A. Purnomo, Sumbangan Indeks Kualitas Udara Wilayah Sebagai Bagian Pencapaian Sustainable Development Goals (Studi Kasus: Kabupaten Magetan), ENVIRO: J. Trop. Environ. Res. 24 (2023), 36-46. https://doi.org/10.20961/enviro.v24i2.70452.

[4] A.H.R. Inaku, C. Novianus, Pengaruh Pencemaran Udara PM 2,5 Dan PM 10 Terhadap Keluhan Pernapasan Anak Di Ruang Terbuka Anak Di DKI Jakarta, Arkesmas 5 (2020), 9–16. https://doi.org/10.22236/arkesmas.v5i2.4990.

[5] A.G. Simandjuntak, Pencemaran Udara, Bul. Limbah. 11 (2007), 34-40.

[6] E. Ertiana, Dampak Pencemaran Udara Terhadap Kesehatan Masyarakat: Literatur Review, J. Ilm. Permas STIKES Kendal. 12 (2022), 287–296.

[7] F. Zhang, H. Cheng, Z. Wang, et al. Fine Particles (PM2.5) at a CAWNET Background Site in Central China: Chemical Compositions, Seasonal Variations and Regional Pollution Events, Atmos. Environ. 86 (2014), 193–202. https://doi.org/10.1016/j.atmosenv.2013.12.008.

[8] R.C.H. Hutauruk, E. Rahmanto, M.C. Pancawati, Variasi Musiman dan Harian PM2.5 di Jakarta Periode 2016–2019, Bul. GAW Bariri. 1 (2020), 20–28. https://doi.org/10.31172/bgb.v1i1.7.

[9] Q. Yang, Q. Yuan, T. Li, et al. The Relationships between PM2.5 and Meteorological Factors in China: Seasonal and Regional Variations, Int. J. Environ. Res. Public Health. 14 (2017), 1510. https://doi.org/10.3390/ijerph14121510.

[10] A.M. Jones, R.M. Harrison, J. Baker, The Wind Speed Dependence of the Concentrations of Airborne Particulate

Matter and NOx, Atmos. Environ. 44 (2010), 1682–1690. https://doi.org/10.1016/j.atmosenv.2010.01.007.

[11] D.M. Westervelt, L.W. Horowitz, V. Naik, et al. Quantifying PM2.5-Meteorology Sensitivities in a Global Climate Model, Atmos. Environ. 142 (2016), 43–56. https://doi.org/10.1016/j.atmosenv.2016.07.040.

[12] M. Puc, B. Bosiacka, Effects of Meteorological Factors and Air Pollution on Urban Pollen Concentrations, Pol. J. Environ. Stud. 20 (2011), 611–618.

[13] M. Tranmer, M. Elliot, Binary Logistic Regression, Cathie Marsh for Census and Survey Research, Paper 20, (2008).

[14] Z.F. Firdaus, D.A. Permadi, Determination of Fine Particulate Matter Episode Periods in Jakarta by Incorporating Meteorological Conditions, J. Presipitasi Media Komun. Pengemb. Tek. Lingkung. 21 (2024), 586–597. https://doi.org/10.14710/presipitasi.v21i2.586-597.

[15] G. Syuhada, A. Akbar, D. Hardiawan, et al. Impacts of Air Pollution on Health and Cost of Illness in Jakarta, Indonesia, Int. J. Environ. Res. Public Health. 20 (2023), 2916. https://doi.org/10.3390/ijerph20042916.

[16] T. Handhayani, An Integrated Analysis of Air Pollution and Meteorological Conditions in Jakarta, Sci. Rep. 13 (2023), 5798. https://doi.org/10.1038/s41598-023-32817-9.

[17] D. Gusnita, N. Cholianawati, Pollutant Concentration and Trajectory Patterns of PM2.5 Including Meteo Factors in Jakarta City, J. Kim. Pend. Kim. 4 (2019), 152. https://doi.org/10.20961/jkpk.v4i3.35028.

[18] Z. Husnina, K. Wangdi, T. Puspita, et al. Profiling Temporal Pattern of Particulate Matter (PM10) and Meteorological Parameters in Jakarta Province during 2020-2021, J. Kesehatan Lingk. 15 (2023), 16–26. https://doi.org/10.20473/jkl.v15i1.2023.16-26.

[19] Z. Chen, D. Chen, C. Zhao, et al. Influence of Meteorological Conditions on PM2.5 Concentrations across China: A Review of Methodology and Mechanism, Environ. Int. 139 (2020), 105558. https://doi.org/10.1016/j.envint.2020.105558.

[20] D. Gujarati, Basic Econometrics (4th edition), The McGraw-Hill, New York, (2004).

[21] R. Arisanti, S. Purnamawati, A. Muslim, Determinants of Greenhouse Gas Emissions in the Transportation Sector in Indonesia: Official Statistics and Big Data Approach, Int. J. Energy Econ. Policy. 14 (2024), 86–97. https://doi.org/10.32479/ijeep.15035.

[22] A. Agresti, An Introduction to Categorical Data Analysis, Wiley, 2007. https://doi.org/10.1002/0470114754.

[23] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Wiley, New York, 2000.

[24] Fractal Analytics Ltd. Comparative Analysis of Classification Techniques: A Fractal White Paper, Fractal Analytics Ltd. (2003).

[25] D.I. Ruspriyanty, A. Sofro, Analysis of Hypertension Disease Using Logistic and Probit Regression, J. Phys.: Conf. Ser. 1108 (2018), 012054. https://doi.org/10.1088/1742-6596/1108/1/012054.

[26] A. Agresti, Categorical Data Analysis, John Willey & Sons, New York, (1990).

[27] D.W. Hosmer, Applied Logistic Regression, John Wiley &Sons, New York, (2000).

[28] R. E. Fitri, E. Setiawan, M. Usman, et al. Analisis Regresi Logistik Biner terhadap Data Indeks Kedalaman Kemiskinan di Indonesia Tahun 2020, J. Siger Mat. 3 (2022), 69–74.

[29] A. Safitri, S. Sudarmin, M. Nusrang, Model Regresi Logistik Biner Pada Tingkat Pengangguran Terbuka Di Provinsi Sulawesi Barat Tahun 2017, VARIANSI: J. Stat. Appl. Teach. Res. 1 (2019), 1. https://doi.org/10.35580/variansiunm9354.

[30] R. Engle, Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, in: Hand book of Econometrics, vol. 2, pp. 775-826, 1984.

[31] A. Nurdini, Cross-Sectional Vs Longitudinal: Pilihan Rancangan Waktu Dalam Penelitian Perumahan Permukiman, Dimensi (J. Tek. Arsit.) 34 (2006), 52–58.

[32] S. Hapsa, R. Khoirudin, Analisis Pertumbuhan Ekonomi D.I. Yogyakarta Tahun 2008-2016, 18 (2018), 142–159