BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0129–0140
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0129-0140

A COMPARISON OF LOGISTIC REGRESSION, MIXED LOGISTIC REGRESSION, AND GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION ON PUBLIC HEALTH DEVELOPMENT IN JAVA

Erwan Setiawan[1*], Muhammad Azis Suprayogi[2], Anang Kurnia[3]

[1]*Program Study of Mathematics Education, Faculty of Teacher Training and Education, Universitas Suryakancana*
Jl. Pasirgede Raya, Bojongherang Kec. Cianjur Kab. Cianjur, 43216, Indonesia

[2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University*
Jl. Meranti Wing 22 level 4, Dramaga, Babakan, Kec. Dramaga, Kab. Bogor, 16680, Indonesia

*Corresponding author's e-mail: *erwan85setiawan@apps.ipb.ac.id*

ABSTRACT

Article History:
Received: 17th April 2024
Revised: 2nd September 2024
Accepted: 2nd September 2024
Published: 13th January 2025

Keywords:
The Public Health Development Index; Logistic Regression; Mixed Logistic Regression; Geographically Weighted Logistic Regression.

The Public Health Development Index (Indeks Pembangunan Kesehatan Masyarakat - IPKM) is a combined parameter that reflects progress in health development and is useful for determining areas that need assistance in improving health development. Through IPKM modeling, factors that significantly influence regional public health development can be discovered. This research aims to find an appropriate model for modeling IPKM and determine the factors that significantly influence public health development. The data used is the 2018 IPKM data collected from 119 cities/regencies in Java. We propose three models namely logistic regression (LR), mixed logistic regression (MLR), and geographically weighted logistic regression (GWLR). The research results show that the MLR is the best model for modeling IPKM in Java based on the AIC value criteria. Based on the MLR model, the factors that have a significant influence on public health development are the egg and milk consumption level and the percentage of the number of doctors per thousand population.

How to cite this article:
E. Setiawan, M. A. Suprayogi, and A. Kurnia., "A COMPARISON OF LOGISTIC REGRESSION, MIXED LOGISTIC REGRESSION, AND GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION ON PUBLIC HEALTH DEVELOPMENT IN JAVA," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0129-0140, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Health development is a planned and sustainable effort to improve the health status of society through various strategies, policies, and interventions. The main goal of health development is to achieve optimal levels of health for the entire population, including increasing access to health services, disease prevention, health promotion, and improving the general quality of life [1]. Measuring health indicators is needed to achieve the goal of sustainable health development. Monitoring health through these indicators allows countries to assess the achievement of international health targets [2]. Therefore, the Indonesian Ministry of Health's Health Research and Development Agency (BALITBANGKES) compiled the Public Health Development Index (*Indeks Pembangunan Kesehatan Masyarakat* - IPKM).

IPKM is a collection of health indicators that can be easily and directly measured to describe health problems. The general principles of indicators used in preparing the IPKM are simple, easy to measure, reliable, and timely. Meanwhile, the IPKM data measures 30 indicators grouped into seven categories, including toddler health, reproductive health, health services, health behavior, non-communicable diseases and their risks, infectious diseases, and environmental health. IPKM serves as a foundation for planning healthcare development programs in districts and cities. It also aids in determining the allocation of healthcare funds from the central government to provinces or districts, and from provinces to regencies or cities [3].

Globally, according to The Legatum Prosperity Index 2017 report, Indonesia's health index is in 101st position out of 149 countries [4]. According to the Global Health Security Index (GHSI) report, Indonesia ranks 13th in global health security among G20 countries. The United States holds the top position with 75.9 points, while Indonesia has 50.4 points on a scale of 100 [5]. With these facts, the IPKM's existence is crucial as a measuring tool for achieving health development in Indonesia.

To enhance the IPKM, we will model it using several variables related to public health. Our research will focus on four independent variables: access to clean water, egg and milk consumption, the ratio of community health centers, and the ratio of doctors. Through IPKM modeling, we hope to identify significant variables that influence the increase in IPKM, allowing policymakers to intervene accordingly. Access to clean water is considered to have a significant influence on IPKM, as it helps prevent infectious diseases, meets nutritional needs, and maintains overall cleanliness [6]. Egg and milk consumption is rich in nutrients, with studies showing that egg consumption can increase good cholesterol levels and that dairy products may benefit bone health and reduce the risk of osteoporosis [7] [8]. Lastly, the ratio of health centers and doctors plays a crucial role in providing easy access to healthcare services, ensuring people receive necessary care, diagnosis, and treatment [9].

Previous research on IPKM has been carried out, namely modeling regencies/cities IPKM in Kalimantan island using probit regression [10]. Research on the public health development index model for East Java province using the geographically weighted logistic regression method [11]. Then research on spatial regression analysis of community health development indices and the Simpson paradox in city or regency in Sumatera [12].

In the 2018 IPKM data, an area is categorized as having health problems (*Daerah Bermasalah Kesehatan* - DBK) if the IPKM value is ≤ 0.6087 (average index) and is classified as non-DBK if the IPKM value is > 0.6087 [3]. Because the IPKM is a binary variable with values 1 or 0 representing non-DBK or DBK, respectively, the appropriate statistical method for modeling is Logistic Regression (LR) [13][14]. Additionally, the IPKM of a Regency/City within a province is likely similar or correlated. This is because health-related policies at the provincial level are often linked to health services at the district/city level, such as budget allocations, health SOPs, and other health programs launched by the provincial government. Consequently, the second model used is Mixed Logistic Regression (MLR), where the provincial variable is incorporated as a random effect for IPKM.

The use of global regression techniques across a large region for the entire study area becomes impractical when there is local interaction with explanatory factors. To tackle this problem, it is recommended to use Geographically Weighted Regression (GWR) models. GWR generates distinct regression models for each specific location [15]. A variant of logistic regression designed for spatial data analysis is Geographically Weighted Logistic Regression (GWLR) [16]. This is based on Waldo Tobler's theory, which states that everything has a relationship with everything else, but things closer to each other will have a stronger relationship than those farther apart [17][18].

In the 2018 IPKM data, there is an imbalance in the response variable. Out of 119 cities/regencies in Java, 101 have non-DBK status, resulting in a proportion of 15% and 85% for DBK and non-DBK status. This imbalance can affect the model's performance. One method to address this issue is Random Over-Sampling Examples (ROSE). This method randomly duplicates samples from the minority class until a relative class balance is achieved. Thus, in this study, there are six model scenarios used to model IPKM, namely LR, MLR, GWLR, LR-Rose, MLR-Rose, and GWLR-Rose. The best model for IPKM modeling was selected based on the AIC value.

2. RESEARCH METHODS

Three models are used in IPKM modeling: Logistic Regression (LR), Mixed Logistic Regression (MLR), and Geographically Weighted Logistic Regression (GWLR). Two scenarios are applied in each model: one without data balancing (LR, MLR, GWLR) and one with data balancing (LR-Rose, MLR-Rose, GWLR-Rose). Analysis of the data was conducted using R software version 4.3.2 and R Studio 2023.09.1, along with some packages, namely readxl, lme4, ggplot, GWmodel, lattice, and VGAM.

2.1 Data

The research data consists of the 2018 IPKM data released by the Ministry of Health of the Republic of Indonesia. This data includes IPKM values for each city/regency in Java. Any city/regency with an IPKM value below the average is referred to as a health problem area (*Daerah Bermasalah Kesehatan* – DBK). The status of a city/regency as DBK or non-DBK is used as the response variable. The predictor variables used to model the response variables are the percentage of households that have access to improved water, expenditure on egg and milk consumption per capita per week (in thousands), percentage of public health centers per thousand population, and percentage of doctors per thousand population. The details of the research variables are explained in **Table 1**.

Table 1. Variables

Variables	Variables Name	Type	Type Description
Y	Public Health Development Index (IPKM)	Binary	1: non-DBK 0: DBK
X1	Percentage of households that have access to improved water	Ratio	
X2	Expenditure on egg and milk consumption per capita per week (in thousands)	Ratio	
X3	Percentage of public health centers per thousand population	Ratio	
X4	Percentage of doctors per thousand population	Ratio	
Z	Province	Nominal	0: DKI Jakarta 1: West Java 2: Central Java 3: Yogyakarta 4: East Java 5: Banten

Source data for X1, X2, X3, and X4 is from the Central Statistics Agency (Badan Pusat Statistik – BPS) of the Province in Java.

2.2 Sampling Class Imbalance Approach

Sampling strategies, specifically over and under-sampling techniques, are widely recognized and successful methods for addressing the issue of class imbalance. To counteract class imbalance by influencing the discrimination process, the ROS algorithm duplicates samples randomly from the minority classes, whereas the RUS technique randomly removes samples from the majority classes until a more balanced distribution of classes is achieved **[19]**. However, both strategies come with certain limitations, including the risk of over-training or over-fitting. Additionally, they may eliminate data that could contribute valuable information for defining the decision boundary. Random Over-Sampling Examples (ROSE) compensates for the class imbalance. This algorithm randomly duplicates samples from the minority class until a relative class balance is achieved [20].

2.3 Logistic Regression (LR)

Logistic Regression is a model of the Generalized Linear Model (GLM) where the response variable is a binary number (0 or 1) and follows a binomial distribution. The formula for the logistic regression model is as follows:

$$\pi_i = \frac{\exp\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}\right)}{1+\exp\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}\right)} \tag{1}$$

where π_i is the probability of $Y_i = 1$ occurring in the i-th observation with $i = 1,2,\dots,n$; β_0 is the intercept of the logistic model; β_j is the parameter value of the j-th independent variable; x_{ij} is the value of the j-th independent variable at the i-th observation; p is the number of independent variables [21].

Parameter estimation uses the Maximum Likelihood Estimation (MLE) method, with the log-likelihood function [22] as follows:

$$\ln L(\beta) = \sum_{i=1}^{n} y_i \ln \pi_i + \sum_{i=1}^{n}(1-y_i)\ln(1-\pi_i) \tag{2}$$

The parameter β which makes the log-likelihood function maximum is obtained when $\frac{\partial \ln L(\beta)}{\partial \beta} = 0$, then the search for β is carried out iteratively using the Newton-Raphson method.

2.4 Mixed Logistic Regression (MLR)

MLR is a form of statistical model that combines elements of logistic regression with mixed models. This random effect is assumed to have a normal distribution. In general, the MLR model is written as follows [23] :

$$\pi_{ij} = \frac{\exp\left(\beta_0+\sum_{i=1}^{p}\beta_i x_{ij} + Z_{ij}u_i\right)}{1+\exp\left(\beta_0+\sum_{i=1}^{p}\beta_i x_i + Z_{ij}u_i\right)} \tag{3}$$

where π_{ij} is the probability $Y_{ij} = 1$ for the i-th observation and j-th group. Z is the covariate matrix for random effects and u_i is the random effect variable. The log-likelihood function for estimating model parameters is as follows [23] :

$$\ln L(\beta,u) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[y_{ij}\ln(\pi_{ij}) + (1-y_{ij})\ln(1-\pi_{ij})\right] \tag{4}$$

The parameters β and u that maximize $\ln L(\beta,u)$ are obtained through $\frac{\partial \ln L(\beta,u)}{\partial \beta} = 0$ and $\frac{\partial \ln L(\beta,u)}{\partial u} = 0$ then the β and u values are obtained through an iterative procedure using the Iteratively Weighted Least Square (IWLS) method.

2.5 Geographically Weighted Logistic Regression (GWLR)

GWLR is a combination of Geographically Weighted Regression (GWR) and Logistic Regression (LR) [24] [25]. GWR and a logistic regression model are combined to form a geographically weighted logistic regression (GWLR) [26]. The GWLR model is formulated as follows:

$$\pi_i = \frac{\exp\left(\beta_0(u_i,v_i)+\sum_{j=1}^{p}\beta_j(u_i,v_i)x_{ij}\right)}{1+\exp\left(\beta_0(u_i,v_i)+\sum_{j=1}^{p}\beta_j(u_i,v_i)x_{ij}\right)} \tag{5}$$

where $\beta_j(u_i, v_i)$ is the coefficient/parameter of the j-th predictor variable on the i-th observation.

A location weighting scheme is required with the assumption that individuals with neighboring locations will have an influence on other individuals, compared to individuals who are far apart [27]. The weighting that will be used is Gaussian fixed kernel function weighting. The Gaussian Fixed Kernel Function is formulated as follows:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2h^2}\right) \tag{6}$$

w_{ij} denotes the weight matrix, d_{ij} denotes the observation distance, and h denotes the bandwidth that affects the observation. To get the optimum h, the Cross Validation (CV) method can be used which is mathematically formulated as follows:

$$CV = \sum_{i=1}^{n}\left(y_i - \hat{y}_{\neq i}(h)\right)^2 \tag{7}$$

Where $\hat{y}_{\neq i}(h)$ is the estimate for y_i where observations at the location (u_i, v_i) are omitted from the estimation process. The optimum bandwidth value is the bandwidth value that causes the minimum CV [28].

The parameter estimation method in the GWLR model is the Maximum Likelihood (MLE) method. The log-likelihood function given the weights is

$$\ln L\left(\beta(u_i, v_i)\right) = \sum_{j=0}^{p}\sum_{i=1}^{n}\left(w_i(u_i, v_i)y_i x_{ij}\right)\beta_j(u_i, v_i) - \sum_{i=1}^{n} w_i(u_i, v_i)\ \ln\left(1 + \exp\left(\sum_{j=0}^{p}\beta_j(u_i, v_i)x_{ij}\right)\right) \tag{8}$$

The parameter $\beta(u_i, v_i)$ that maximizes $\ln L\left(\beta(u_i, v_i)\right)$ is obtained through $\frac{\partial \ln L(\beta(u_i,v_i))}{\partial \beta_j(u_i,v_i)} = 0$, which then the value of $\beta(u_i, v_i)$ is obtained through an iterative procedure using the Iteratively Weighted Least Square (IWLS) algorithm.

2.6 Model Selection

The method used to select the best model is Akaike's Information Criterion (AIC) which is defined as follows:

$$AIC = -2 \cdot \log(L) + 2k \tag{9}$$

where $\log(L)$ is the log-likelihood of the model to the data and k is the number of free parameters in the model. The best model is the model with the smallest AIC value [29].

2.7 Analysis Procedure

The data analysis procedure is followed in **Figure 1**.

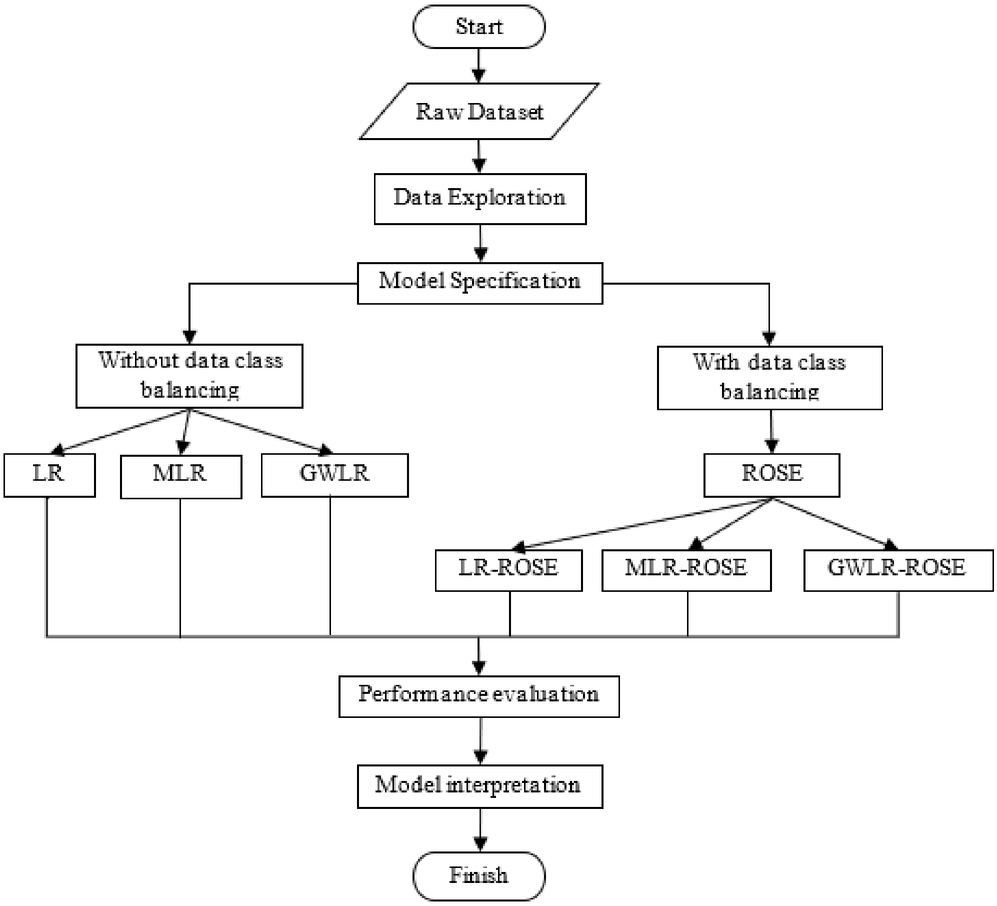


Figure 1. Research flow

According to **Figure 1**, the research flow involves two distinct scenarios: one where modeling is conducted without data class balancing, and another where modeling is carried out with data class balancing. In both scenarios, three models, LR, MLR, and GWLR, are used. The data class balancing technique applied is ROSE.

3. RESULTS AND DISCUSSION

3.1 Data Exploration

The number of research data is 119 observations from districts and cities on the island of Java in 2018. class comparison on the response variable, namely the IPKM, is 85% for IPKM equal to or greater than the average and 15% for IPKM less than the national average as shown in **Figure 2**.

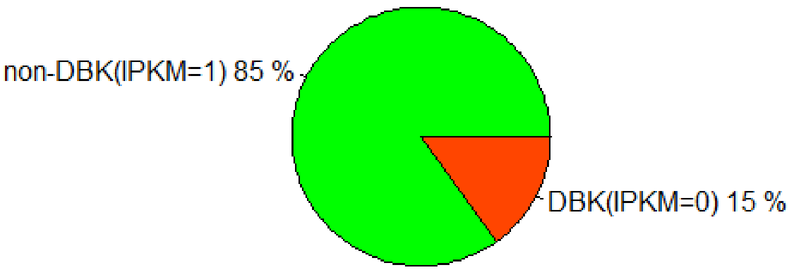


Figure 2. Percentage of IPKM value in Java Island

The boxplot in **Figure 3** shows the distribution of predictor variables on response variables. The percentage of households with access to adequate water (X1) has a higher median at IPKM=1 (non-DBK), indicating that as more households have access to improved water, the IPKM level is higher. Similarly, the variable Expenditure on egg and milk consumption per capita per week (in thousands) (X2) has a higher median boxplot at a high IPKM level, suggesting that higher consumption of egg and milk leads to a higher IPKM level. In contrast, the variable number of community health centers per thousand population (X3) shows similar boxplots for both high and low IPKM levels, with low values and a few outlier values. The variable number of doctors per thousand population (X4) has low values at both high and low IPKM levels, but the number of doctors per population is greater at high IPKM levels. This indicates that a higher number of doctors per population leads to a higher IPKM level. Additionally, the number of outliers in variable X4 is quite large at both IPKM levels.

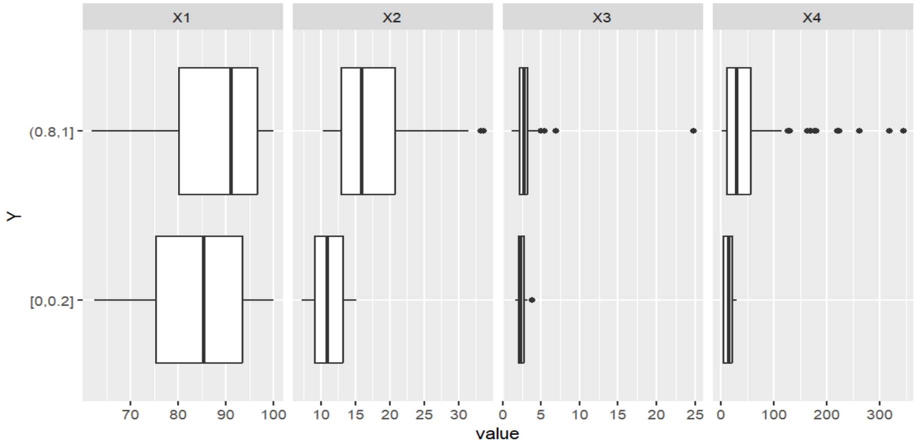


Figure 3. Distribution of Variables Based on IPKM

3.2 Data Analysis

This study utilizes three main models: LR, MLR, and GWLR. LR is employed to identify the factors influencing the IPKM value, while MLR is used to determine these factors while also taking into account random effects at the provincial level. The GWLR model is used to examine the spatial influences on the IPKM value of a region. Within the GWLR model, the adaptive Gaussian kernel function is used. This model involves specifying the geographic coordinates (latitude and longitude) for each observation area, which take the form of districts or cities. A total of 119 cities/regencies were observed. The IPKM values for each district and city in 2018 are depicted in **Figure 4** on the map.

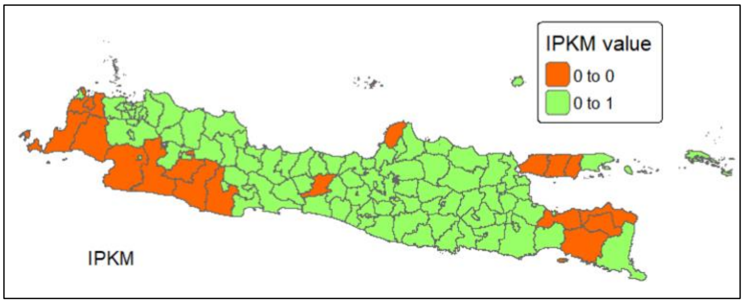


Figure 4. IPKM Distribution by Maps in the Year 2018

Based on 2018 IPKM data on the island of Java, several districts still have low IPKM levels spread across four provinces, namely West Java Province, Central Java Province, East Java Province, and Banten Province. In West Java Province there are Sukabumi Regency, Cianjur Regency, Bandung Regency, Garut Regency, Tasikmalaya Regency. In Central Java Province, there are Banjarnegara Regency and Jepara Regency. For East Java Province there are Bondowoso Regency, Situbondo Regency, Probolinggo Regency, Bangkalan Regency, Sampang Regency and Pamekasan Regency. Meanwhile, in Banten Province, there are Pandeglang Regency, Lebak Regency, Serang Regency, and Serang City.

The response data from the IPKM has 2 categories: a value equal to or above the IPKM average (IPKM=1) and a value below the IPKM average (IPKM=0). The distribution is unbalanced, with 85% for IPKM=1 and only 15% for IPKM=0, as depicted in Figure 2. Therefore, this study considers 2 scenarios. To address this imbalance, the minority data is duplicated at random through oversampling. The quantity of duplication is selected to optimize the regression value. The oversampling of the minority class utilizes the Random Over Sampling Examples (ROSE) package. Figure 5 illustrates the comparison of the new data following the oversampling procedure.

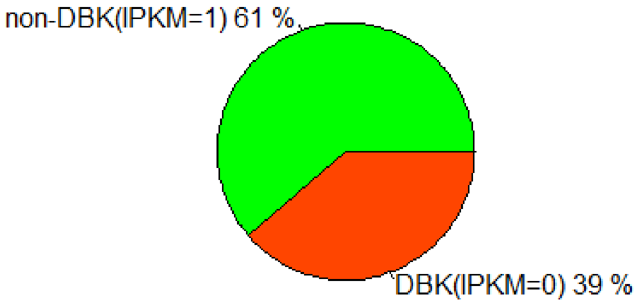


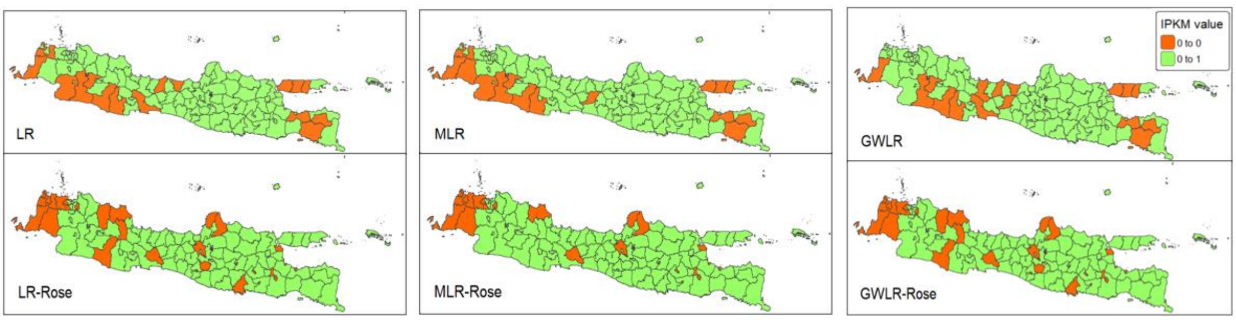
Figure 5. Percentage of IPKM value after Over Sampling

In this study, IPKM modeling uses three models with two scenarios for each model. The selection of the most appropriate model is determined using the AIC value criteria. The model with the smallest AIC value is the best in modeling IPKM. The AIC value for each model can be seen in Table 2.

Table 2. The AIC Value for Each Model

Models	AIC
LR	51.281
MLR	51.2
GWLR	54.052
LR-Rose	90.317
MLR-Rose	83.8
GWLR-Rose	90.317

From looking at Table 2, scenario 1 reveals that there is minimal difference in the AIC value between the logistic regression model (LR) and the mixed logistic regression model (MLR). However, the GWLR model exhibits a lower goodness of fit value with an AIC value greater than that of the LR and MLR models. This suggests that the inclusion of spatial factors in the GWLR model does not offer additional information in interpreting the IPKM index for a region compared to the LR and MLR models. In scenario 2, where the LR, MLR, and GWLR models are balanced by oversampling, it is evident that the AIC value is higher than the model without oversampling. Therefore, a model without data balancing is preferable to a model with data balancing. Considering the AIC value, the MLR model emerges as the best model.



(a) **(b)** **(c)**

Figure 6. Plot between Predicted Values from Models and the Actual Values
(a) Model LR and LR-Rose, (b) Model MLR and MLR-Rose, (c) Model GWLR and GWLR-Rose

The predicted values from each model are visualized on a geographic map according to **Figure 6**. The best model is the color pattern on the map of the island of Java which is closest to the color pattern of the map of the island of Java based on actual data as in **Figure 4**.

The MLR model best replicates the actual data image, with discrepancies observed in specific areas, such as Jepara Regency, Bondowoso Regency, Situbondo Regency, Probolinggo Regency, Bandung Regency, and West Bandung Regency. Following closely is the LR model, albeit with a slightly lower similarity to the actual data image compared to the MLR model. The GWLR model ranks third, reflecting a moderate similarity to the actual data. The figure indicates that incorporating spatial factors through the GWLR model does not bring additional insights into the context of IPKM above or below the average. Additionally, the introduction of data balancing to these models results in substantial deviation from the actual data, leading to the conclusion that the data balancing process does not enhance model performance, but instead yields inferior outcomes. It is, therefore, determined that the analysis and interpretation shall be focused solely on the MLR model, given the adequacy of using MLR based on the above results.

3.3 Analysis and Interpretation of Mixed Logistic Regression Model (MLR)

Multicollinearity checks on predictor variables involving a combination of numerical and categorical predictors can be performed using Variance Inflation Factors (VIF). The likelihood of multicollinearity among variables increases as the tolerance decreases. A VIF value of 1 suggests that the independent variables are not correlated. If the VIF is between 1 and 5, it indicates a moderate correlation among variables. The critical range for VIF is 5 to 10, signifying highly correlated variables. When VIF is equal to or exceeds 5 to 10, multicollinearity is present among predictors in the regression model, and a VIF greater than 10 suggests that the regression coefficients are weakly estimated due to multicollinearity **[30]**, **[31]**. The VIF can be seen in **Table 3**.

Table 3. VIF of Predictor Variables

Variable	VIF
X1	1.243
X2	1.851
X3	1.370
X4	1.241

As shown in **Table 3**, the VIF for all predictor variables is below 5, indicating the absence of multicollinearity. Therefore, it is feasible to proceed with MLR modeling. **Table 5** provides the parameter estimates and results of the partial MLR parameter testing.

Table 5. Estimate of Mixed Logistic Regression Model Parameters

Parameters	Estimate	Std Error	z value	p-value	Odds Ratio
β_0	-34.82903	13.21100	-2.636	0.00838	
β_1	0.13716	0.08813	1.556	0.11963	1.14701
β_2	1.40637	0.47032	2.990	0.00279	4.08111
β_3	1.23457	0.91681	1.347	0.17811	3.43690
β_4	0.17872	0.08044	2.222	0.02630	1.19569

According to the results presented in **Table 5**, the Wald test indicates that the predictor variables with significant impact on the IPKM index at a 5% significance level are the expenditure on egg and milk consumption per capita per week (in thousands) (X2) and the number of doctors per population (X4).

The coefficient of expenditure on egg and milk consumption per capita per week (in thousands) (X2) is positive (1.40637), meaning that the higher the expenditure on egg and milk consumption per capita per week, the higher the estimated probability of high level of IPKM when other variables are held constant. The odds ratio value of 4.081114 means that for every increase in expenditure on egg and milk consumption per capita per week (in thousands), the estimated odds of high index IPKM become 408.1114% of the estimated odds before the expenditure increases.

The coefficient of the percentage of doctors per thousand population (X4) is positive (0.17872), meaning that the higher the percentage of doctors per thousand population, the higher the estimated probability of a high level of IPKM (IPKM=1) when other variables are held constant. The odds ratio value of 1.195686 means that for every 1% increase in the percentage of doctors per thousand population, the estimated odds of high index IPKM become 119.5686% of the estimated odds before the percentage of doctors per thousand population increases.

The influence of other variables, such as the percentage of households with access to improved water (X1) and the percentage of public health centers per thousand population (X3), does not exhibit significant effects on the low-level or high-level index of IPKM at a significance level of $\alpha = 5\%$. The lack of significant impact of the percentage of households with access to improved water in this study can be attributed to the similarity in its distributions across both the low-level and high-level index of IPKM as illustrated in **Figure 3**. Similarly, the non-significant effects of the percentage of public health centers per thousand population in this study can be ascribed to the similar proportion of each variable within the low-level and high-level index of IPKM, as depicted in **Figure 3**. This interpretation elucidates the findings related to these variables.

The positive coefficient (0.13716) indicates that as the percentage of households with access to improved water increases, the estimated probability of a high-level index of IPKM decreases when holding other variables constant. With an odds ratio of 1.14701, a one-percent increase in households with access to improved water leads to the estimated odds of a high-level index of IPKM 114.701% of the estimated odds before the increase in access to improved water.

The coefficient of the percentage of public health centers per thousand population is positive (1.23457), meaning that the higher the percentage of public health centers per thousand population, the smaller the estimated probability of a high-level index of IPKM when other variables are held constant. The odds ratio value of 3.43690 means that for every one-percent increase in public health centers per thousand population, the estimated odds of the high-level index of IPKM become 343.690% of the odds before the percentage of public health centers per thousand population increases.

The estimated parameter values of the random effects can be seen in **Table 6**, while their significance is shown in **Figure 7**.

Table 6. Estimation of The Random Effect

Province	Estimate
Jakarta	8.054×10^{-15}
West Java	0.53982
Central Java	2.74556
Yogyakarta	0.01526
East Java	-0.55195
Banten	-3.03413

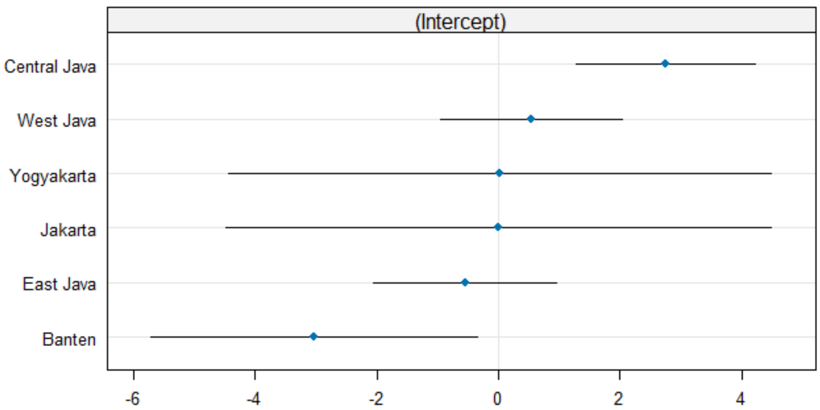


Figure 7. Confidence Interval of the Random Effect

The random effects of West Java, Yogyakarta, Jakarta, and East Java are not found to be significant in **Figure 7**, whereas Central Java and Banten show significant random effects. A significant random effect indicates a real influence on the outcome that cannot be fully explained by the fixed effects in the model.

The formula for the basic MLR model is represented in **Equation (3)**. Specifically, the MLR model in the province of West Java was developed as described by

$$\pi_i = \frac{\exp(-34.82903 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4} + 0.53982)}{1 + \exp(-34.82903 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4} + 0.53982)}$$

$$= \frac{\exp(-34.28921 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4})}{1 + \exp(-34.28921 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4})}$$

If a 1-unit change is made in only variable X2, then the likelihood of IPKM being equal to 1 in West Java is

$$\pi_i = \frac{\exp(-34.28921 + 1.40637(1))}{1 + \exp(-34.28921 + 1.40637(1))} = 5.23798 \times 10^{-15}$$

Meanwhile, the MLR model located in Central Java is

$$\pi_i = \frac{\exp(-34.82903 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4} + 2.74556)}{1 + \exp(-34.82903 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4} + 2.74556)}$$

$$= \frac{\exp(-31.54365 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4})}{1 + \exp(-31.54365 + 0.13716x_{i1} + 1.40637x_{i2} + 1.23457x_{i3} + 0.17872x_{i4})}$$

If only X2 undergoes a 1-unit intervention, then the likelihood of IPKM equaling 1 in Central Java will be impacted.

$$\pi_i = \frac{\exp(-32.08347 + 1.40637(1))}{1 + \exp(-32.08347 + 1.40637(1))} = 4.7545 \times 10^{-14}$$

Thus, the Odd ratio is

$$\frac{4.7545 \times 10^{-14}}{5.23798 \times 10^{-15}} = 9.08 \sim 9$$

When the variable X2 is intervened, the likelihood of IPKM =1 in Central Java is 9 times greater than in West Java.

4. CONCLUSIONS

In this study, there was no difference between the logistic regression (LR) model and the mixed logistic regression (MLR) model in the context of the index level of IPKM based on the values of AIC and the ratio of deviance and degrees of freedom (df). The GWLR model is less than the LR and MLR models shows that the addition of spatial factors through the GWLR model does not provide additional information in the context of the index level of IPKM. The research results show that the MLR is the best model for modeling IPKM in Java based on the AIC value criteria. The factors significantly affecting the index level of IPKM based on the mixed logistic regression model are expenditure on egg and milk consumption per capita per week (in thousands) and percentage of doctors per thousand population, while other variables, namely the percentage of households that have access to improved water and percentage of public health centers per thousand population have no significant effects.

REFERENCES

[1] R. Labonte and G. Laverack, *Capacity Building in Health Promotion, Part I: For Whom? And for What Purpose?* vol. 11:

Crit. Public Health, 2001.

[2] World Health Organization (WHO), "World Health Statistics 2021: Monitoring Health for SDGs," 2021.

[3] D. H. Tjandrarini and dkk, *Indeks Pembangunan Kesehatan Masyarakat 2018*. Jakarta: Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan, 2019.

[4] Y. Debora, "Indeks Kesehatan Indonesia Masih Sangat Rendah," *tirto.id*, 2017. https://tirto.id/indeks-kesehatan-indonesia-masih-sangat-rendah-cBRn (accessed Nov. 28, 2023).

[5] A. M. H. Putri, "Perhatian! Indeks Ketahanan Kesehatan RI Masih Jauh di Bawah," *CNBC Indonesia, 2023*, 2023.

[6] World Health Organization (WHO), "Water, Sanitation, Hygiene, and Health," 2022.

[7] J. Y. Shin, P. Xun, Y. Nakamura, and K. He, "Egg consumption in relation to risk of cardiovascular disease and diabetes: a systematic review and meta-analysis.," *Am. J. Clin. Nutr.*, vol. 98, no. 1, pp. 146–159, 2013, doi: 10.3945/ajcn.112.051318.

[8] D. Feskanich, W. C. Willett, and G. A. Colditz, "Calcium, vitamin D, milk consumption, and hip fractures: a prospective study among postmenopausal women," *Am. J. Clin. Nutr.*, vol. 77, no. 2, pp. 504–511, 2003, doi: 10.1093/ajcn/77.2.504.

[9] World Health Organization (WHO), "Everybody's business: Strengthening health systems to improve health outcomes: WHO's framework for action.," *World Health Organization (WHO)*, 2016.

[10] M. Fathurahman, "Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Pulau Kalimantan Menggunakan Pendekatan Regresi Probit," *J. VARIAN*, vol. 2, no. 2, pp. 47–54, 1970, doi: 10.30812/varian.v2i2.382.

[11] Q. S. Wardhani, S. S. Handajani, and I. Susanto, "Masyarakat Jawa Timur dengan metode," *J. Apl. Stat. dan Komputasi*, vol. 14, no. 2, pp. 1–12, 2022.

[12] U. K. Krismayanto and E. Pasaribu, "Analisis Regresi Spasial Indeks Pembangunan Kesehatan Masyarakat dan Paradoks Simpson Kabupaten/Kota di Pulau Sumatera Tahun 2018," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 1037–1052, 2022, doi: 10.34123/semnasoffstat.v2022i1.1330.

[13] M. A. Santika and Y. Karyana, "Analisis Regresi Logistik Biner dengan Efek Interaksi untuk Memodelkan Angka Fertilitas Total di Jawa Barat," in *Bandung Conf. Ser. Stat.*, 2022, p. vol. 2, no. 2, pp. 142–151. doi: doi: 10.29313/bcss.v2i2.3555.

[14] I. Haq, M. N. Aidi, A. Kurnia, and E. Efriwati, "A Comparison of Logistic Regression and Geographically Weighted Logistic Regression (Gwlr) on Covid-19 Data in West Sumatra," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 3, pp. 1749–1760, 2023, doi: 10.30598/barekengvol17iss3pp1749-1760.

[15] and S. F. M. Rodrigues, J. de la Riva, "Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression," *Appl. Geogr.*, vol. 48, pp. 52–63, 2014, doi: 10.1016/j.apgeog.2014.01.011.

[16] M. Fathurahman, Purhadi, Sutikno, and V. Ratnasari, "Pemodelan Geographically Weighted Logistic Regression pada Indeks Pemodelan Geographically Weighted Logistic Regression pada Indeks Pembangunan Kesehatan Masyarakat di Provinsi Papua," in *in Prosiding Seminar Nasional MIPA 2016*, 2016, pp. 34–42.

[17] A. Getis, *Waldo Tobler (1931-2018): Analytical Cartographer and Regional Scientist.* 2020.

[18] A. T. Murray et al., "Overview of Contributions in Geographical Analysis : Waldo Tobler," *Geogr. Anal.*, vol. 52, pp. 480–493, 2020.

[19] P. Vuttipittayamongkol and E. Elyan, "Neighbourhood-based undersampling approach for handling imbalanced and overlapped data," *Inf. Sci. (Ny).*, vol. 509, no. October, pp. 47–70, 2020, doi: 10.1016/j.ins.2019.08.062.

[20] E. Rendón, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez, "Data sampling methods to deal with the big data multi-class imbalance problem," *Appl. Sci.*, vol. 10, no. 4, 2020, doi: 10.3390/app10041276.

[21] P. McCullagh and J. A. Nelder, *Generalized Linear Models second edition*. New York: Chapman & Hall, 1989.

[22] A. Agresti, *Foundations of Linear and Generalized Linear Models*. New Jersey: John Wiley & Sons, Inc, 2015.

[23] C. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*. New Jersey: John Wiley & Sons, Inc, 2000. doi: 10.1002/0471722073.

[24] P. M. Atkinson, S. E. German, D. A. Sear, and M. J. Clark, "Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression Biometrika," *Geogr. Anal.*, vol. 35, pp. 59–82, 2003.

[25] M. Fathurahman, Purhadi, Sutikno, and V. Ratnasari, "Geographically Weighted Multivariate Logistic Regression Model and Its Application," *Abstr. Appl. Anal.*, vol. 2020, 2020, doi: 10.1155/2020/8353481.

[26] M. Fathurahman, Purhadi, Sutikno, and V. Ratnasari, "Hypothesis testing of Geographically weighted bivariate logistic regression," *J. Phys. Conf. Ser.*, vol. 1417, no. 1, 2019, doi: 10.1088/1742-6596/1417/1/012008.

[27] C. Chasco, I. Garcia, and J. Vicens, "Modeling Spatial Variations in Household Disposable Income With Geographically Weighted Regression," *Munich Pers. RePEc Arch.*, 2007, [Online]. Available: https://mpra.ub.uni-muenchen.de/1682/%0D

[28] W. S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," *J. Am. Stat. Assoc.*, vol. 74, no. 368, 1979, doi: https://doi.org/10.1080/01621459.1979.10481038.

[29] A. Hirotugu, "A New Look at The Statistical Model Identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.

[30] D. A. Belsley, *Conditioning diagnostics: Collinearity and weak data in regression*. New York, USA: John Wiley & Sons, Inc, 1991.

[31] N. Shrestha, "Detecting Multicollinearity in Regression Analysis," *Am. J. Appl. Math. Stat.*, vol. 8, no. 2, pp. 39–42, 2020, doi: 10.12691/ajams-8-2-1.