Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Comparison of distance-based spatial weight matrix in modeling Internet signal strengths in Tasikmalaya regency using logistic spatial autoregressive model

Yuhana Notonegoro[a], Yudhie Andriyana[b*] and Budi Nurani Ruchjana[c]

[a]*Post-Graduate Program in Applied Statistics, Faculty of Mathematics and Natural Sciences,Universitas Padjadjaran, Sumedang 45363, Indonesia*
[b]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia*
[c]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang 45363, Indonesia*

C H R O N I C L E

Article history:
Received: November 10, 2023
Received in revised format: November 20, 2023
Accepted: December 15, 2023
Available online: December 15, 2023

Keywords:
Signal Strength
Distance-Based Spatial Weight Matrix
Optimization
Logistic SAR Model
Bayesian MCMC
Confusion Matrix

A B S T R A C T

To ensure that national development objectives in rural areas are achieved evenly and sustainably, the Government of Indonesia applies the principles of Village Sustainable Development Goals (SDGs), which are derivative programs of SDGs. One of the indicators in measuring the progress and independence of villages in Indonesia is the availability of cellular phone signal access. Cellular phone signals have a vital role because most internet users in Indonesia rely on mobile data connections from cellular operators. However, the signal emitted by a provider tower has a limited range. According to the data of the Developing Villages Index in 2022, Tasikmalaya Regency is one of the regencies with the highest number of villages that have weak signal strength in West Java Province, Indonesia. To examine the effect of distance and height difference between the placement of the nearest provider tower and the location of the Village Office on the internet signal strength category in Tasikmalaya Regency, Logistic Spatial Autoregressive modeling is needed. In this study, the Bayesian Markov-Chain Monte Carlo estimation method was used, because it has advantages in flexibility and computational efficiency. In spatial modeling, there is a spatial weight matrix determined by the researcher's understanding of the observed phenomenon. The variable observed in this study is signal strength, which has an orientation at a distance. However, there are several types of distance-based spatial weight matrices, such as K-nearest neighbor, radial distance, power distance, and exponential distance. To determine the most suitable distance-based spatial weight matrix in internet signal strength modeling, the four (4) weight matrices were compared based on the goodness of fit measure models, calculated from the confusion matrix. The results of the analysis showed that the radial distance weight matrix with a threshold distance of d = 1.7km is the most suitable use of distance-based spatial weight matrix in internet signal modeling in Tasikmalaya Regency. The weight matrix exerted a positive spatial autocorrelation effect of 57.141%. In addition, the height difference factor between the location of the provider tower with the location of the village office has a greater effect than the horizontal distance.

© 2024 by the authors; licensee Growing Science, Canada.

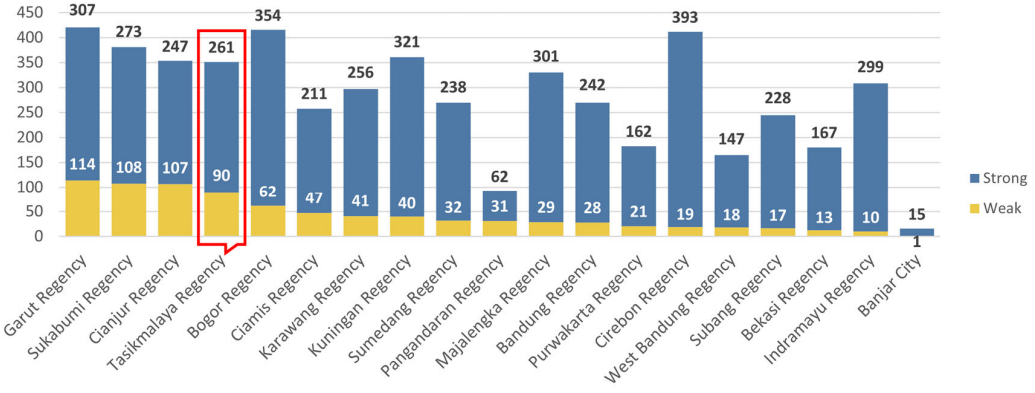
1. Introduction

Most of the mainland of Indonesia, which is 89.81%, is a rural area (Ministry of Home Affairs Republic of Indonesia, 2022). To ensure the national development objectives in rural areas are achieved evenly and sustainably, in accordance with Presidential Regulation Number 59 of 2017 concerning the Implementation of Achieving Sustainable Development Goals (Iskandar, 2020), the Indonesian Government applies the principles of Village Sustainable Development Goals (SDGs), a program derived from SDGs issued by the United Nations. Indonesia has a Developing Villages Index (IDM) as one of the indicators to measure the villages' progress and independence. One of the measurement indicators is the availability of cellular

* Corresponding author.
E-mail address: y.andriyana@unpad.ac.id (Y. Andriyana)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)
© 2024 by the authors; licensee Growing Science, Canada.
doi: 10.5267/j.ijdns.2023.12.016

phone signal access. This indicator is stated in the regulation of the Minister of Villages, Development of Disadvantaged Regions, and Transmigration Number 2 of 2016 concerning the Developing Villages Index (2016), on the index of social resilience of the settlement dimension, access to information, and communication. The availability of cellular phone signals can facilitate access to communication and disclosure of public information. Cellular phone signals have a vital role, especially in Indonesia. The reason is that 77.64% of Interconnected Network (internet) users in Indonesia access internet connections through mobile data from cellular operators (Indonesia Internet Service Provider Association, 2022). Accessing an internet connection in this way has many benefits and advantages; easier for users to access anytime and anywhere, easier to purchase data packages, cheaper, and has many attractive promos. Even so, the signal emitted by a provider tower has a limited coverage area (Lee, 2006). The farther the distance reached by the signal from the center of the transmitter, the decrease in the strength of the signal received. Therefore, the placement of a signal transmitter location can affect the capacity and quality of the network.



Source: IDM Data, Community and Village Empowerment Office of West Java Province

Fig. 1. The Number of Villages Based on The Cellular Phone Signal Strength by Regency in West Java Province, Indonesia in 2022

According to IDM data (2022), out of a total of 351 villages in Tasikmalaya Regency, there are 90 villages (25.64%) that have weak signal strength. Based on Figure 1, Tasikmalaya Regency is among the five (5) districts with the highest number of villages with weak signals in West Java Province, Indonesia, with 1.69% of the total 5,312 villages in West Java. The signal strength characteristic is a radius that spreads in all directions. If an area has a strong signal, the farther the distance of the surrounding areas, the more decreased the signal strength possessed by the other areas. This shows that signal strength tends to have a spatial effect. Therefore, to examine the effect of distance and height difference between the placement of the nearest provider tower and the location of the village office on the internet signal strength category in Tasikmalaya Regency, Logical Spatial Autoregressive modeling was carried out. The estimation method used in this study is the Bayesian Markov-Chain Monte Carlo (Bayesian MCMC) (Krisztin & Piribauer, 2020). The method proposed by Krisztin and Piribauer has the advantages of flexibility in modeling, as well as being computationally easier and more efficient. In their research, they applied the method to model the Foreign Direct Investment (FDI) decision spatially in Europe from multinational companies. The spatial weight matrix they used is the 8-nearest neighbors weight matrix.

In theory, the spatial weight matrix is fixed, determined by the assumptions of researchers based on the understanding of the observed phenomenon. However, it is not uncommon for researchers not to have enough information to determine the type of spatial weight matrix, causing researchers to refer to Tobler's law (1970) which states that everything is related to everything else, but near things are more related than distant things. This law is often translated to the use of a contiguity-based spatial weight matrix as the right decision in their spatial modeling. In contrast, signal strength very clearly has an orientation at a distance. However, there are many types of distance-based spatial weight matrices, namely K-Nearest Neighbor (K-NN), radial distance, power/inverse distance, and exponential distance (Zhou & Lin, 2008). In this study, to determine the most suitable distance-based spatial weight matrix in describing the phenomenon of internet signal strength in Logistic SAR modeling in Tasikmalaya Regency, the four (4) weight matrices are compared based on the goodness of fit measure model, calculated from the confusion matrix.

2. Method

2.1 Data Source

The data used in this study is secondary data obtained from IDM raw data by the Ministry of Villages, Development of Disadvantaged Regions, and Transmigration Republic of Indonesia through the Community and Village Empowerment Office of West Java Province, Indonesia and OpenCellID website. The data is cross-sectional data consisting of one (1) response

variable and two (2) predictor variables with an observation unit of 351 village offices in Tasikmalaya Regency ($N = 351$) in 2022. Fig. 2 and Table 1 below show a map of village office location points along with the research variables used in this study:

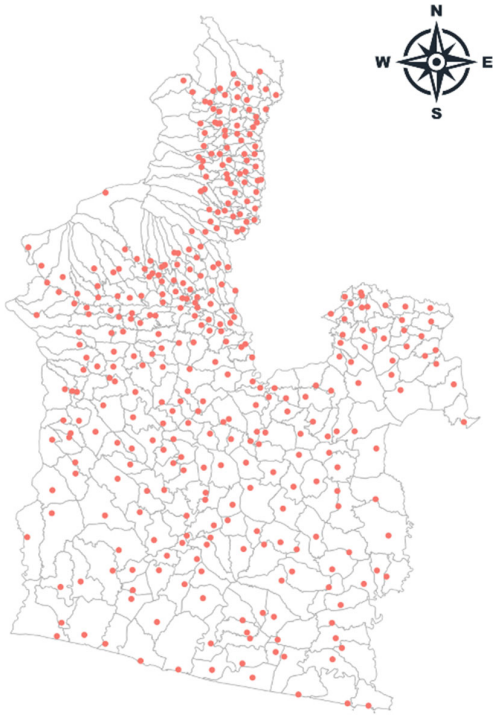


Fig. 2. Map of Administrative and Village Office Location Points of Tasikmalaya Regency

Table 1
Research Variables

Variable Name	Unit	Source
Cellular phone/handphone signal strength in the village (Y)	1 = Strong, and 0 = Weak	Community and Village Empowerment Office of West Java Province
Distance of the nearest provider tower to the village office/center of the village/village settlement (X_1)	Meters	Processed results between data from Community and Village Empowerment Office of West Java Province and OpenCellID
Height difference between the nearest provider tower placement and village office location (X_2)	Meters	Processed results between data from Community and Village Empowerment Office of West Java Province and OpenCellID

2.2 Logistic Spatial Autoregressive Model

The spatial logistic regression model in this study is a combination of the logistic regression model with SAR. The form of the Logistic SAR model equation according to Pinkse & Slade (1998) is as follows:

$$\begin{aligned} \boldsymbol{Y}^* &= \rho\mathbf{W}\boldsymbol{Y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{Y}^*(\mathbf{I}_N - \rho\mathbf{W}) &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{Y}^* &= (\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon} \ , \end{aligned} \tag{1}$$

where ρ is the lag coefficient parameter of the response variable which indicates the strength of spatial autocorrelation; $\mathbf{W}$ is the spatial weight matrix of size ($N \times N$); $\mathbf{X} = [\mathbf{1}, \boldsymbol{X}_1, \boldsymbol{X}_2]$ is the predictor variable matrix of size ($N \times (p + 1)$); $\boldsymbol{\beta}$ is regression parameter vector of size ($(p + 1) \times 1$); $\boldsymbol{\varepsilon}$ is the random error vector model of size ($N \times 1$) that is independently and identically Gaussian distributed, with mean zero (0) and variance σ^2 ($\boldsymbol{\varepsilon} \sim iidN(\mathbf{0}, \sigma^2\mathbf{I}_N)$); N is the number of observations in the population; p is the number of predictor variables; and $\boldsymbol{Y}^*$ is the vector of latent variables of size ($N \times 1$) measured through $logit(\pi(\mathbf{X}))$. In the logistic regression model, $\mathbf{X}\boldsymbol{\beta}$ is a linear predictor, so the logistic SAR model in Eq. (1) has the complete form as follows:

$$\boldsymbol{Y}^* = logit\big(\pi(\mathbf{X})\big) = (\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} \ ,$$

with

$$\pi(\mathbf{X}) = \frac{\exp((\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta})}{1 + \exp((\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta})} \tag{2}$$

The latent variable vector $\boldsymbol{Y}^*$ has two (2) categories defined as:

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* = logit\big(\pi(X_i)\big) > 0 \\ 0 & \text{if } Y_i^* = logit\big(\pi(X_i)\big) \leq 0 \end{cases} \tag{3}$$

In general, the parameters of the Logistic SAR model can be estimated using several approaches. In this study, the Bayesian MCMC estimation method was used because it has the advantages of flexibility in modeling, as well as being computationally easier and more efficient (Krisztin & Piribauer, 2020).

2.3 Bayesian Markov-Chain Monte Carlo

Based on the logit model, the probability model as a function of log-odds μ_i:

$$P_i = P(Y_i = 1) = \frac{\exp(\mu_i)}{1 + \exp(\mu_i)} \tag{4}$$

In a standard logit model, μ_i is usually specified as a linear combination of a matrix of explanatory variables and corresponding vectors of unobserved intercept and slope parameters. The core part of the Logistic SAR model in this approach as in Eq. (1), which is defined as follows:

$$\boldsymbol{\mu} \ = \rho\mathbf{W}\boldsymbol{\mu} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{A}^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \ , \tag{5}$$

with $\mathbf{A}^{-1} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}$,

where vector $\boldsymbol{\mu}$ is a latent variable $\boldsymbol{Y}^*$, which size ($N \times 1$) with log-odds $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)^{\mathrm{T}}$ also depends on the characteristics of other regions in the sample. Spatial dependencies are introduced by a spatial multiplier that states that $(\mathbf{I} - \rho\mathbf{W})^{-1} = \sum_{r=0}^{\infty} \rho^r \mathbf{W}^r$. The random error in the Logistic SAR model consists of two different components: first, heteroscedasticity random errors that arise from the model in Eq. (4), and second, random error $\boldsymbol{\varepsilon}$ which is normally distributed with variance σ^2. Spatial dependencies in random error are included in the second random error component. Like the variance with Spatial Probit Model, variant σ^2 is limited, this is to identify the correct logistic random error. The likelihood function of the conditional parameter vector ($N \times 1$) observed in option Y can be written as follows:

$$\prod_{i=1}^{N} \frac{\exp(\mu_i)^{Y_i}}{1 + \exp(\mu_i)} \tag{6}$$

the likelihood function of the i-th observation depends on probabilities $P(Y_i = 1)$ and $P(Y_i = 0)$. This results in the problem that the log-likelihood function is not linear, this can greatly complicate the process of estimating and inferring parameters. To efficiently handle this complex actual logit framework, an estimation strategy with latent Pólya-Gamma variables is used (Polson, Scott, & Windle, 2013). With this latent variable, it is possible to reconstitute the conditional posterior distribution from parameter $\boldsymbol{\beta}$ present in the logit framework to a Gaussian distribution. The introduction of the Pólya-Gamma variable can greatly facilitate the Bayesian estimation process for the specification of Binomial type models. The following identities are given based on the results of Polson, et al. (2013):

$$\frac{(\exp \mu_i)^a}{(1 + \exp \mu_i)^b} = 2^{-b} \exp(k_i \, \mu_i) \int_0^{\infty} \left(\exp \frac{-\omega_i \, \mu_i^2}{2} \right) p(\omega_i) d\omega_i \ . \\ a, \mu_i \in \mathbb{R}, b \in \mathbb{R}^+ \mathcal{PG}, \tag{7}$$

where $k_i = \alpha - (b/2)$, and ω_i are random variables with a Pólya-Gamma distribution with scale parameter b and location parameter zero (0), $p(\omega_i) \sim \mathcal{PG}(b, 0)$. The integral identity of Eq. (7) does not depend on numerical estimates of the Binomial distribution and can be estimated by sampling from the Pólya-Gamma conditional posterior distribution. Based on the results of Polson, et al. (2013) and Windle, et al. (2014), the conditional posterior $\boldsymbol{\omega} = (\omega_1, ..., \omega_N)^{\mathrm{T}}$ also takes the form of a Pólya-Gamma distribution:

$$p(\boldsymbol{\omega}|\boldsymbol{\beta},\rho,\boldsymbol{Y}) = \mathcal{PG}(1, \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{A}^{-1}\boldsymbol{\varepsilon}) \ , \tag{8}$$

A computationally efficient sampling algorithm for the Pólya-Gamma conditional posterior distribution is contained in Eq. (8) (Polson, Scott, & Windle, 2013; Windle, Polson, & Scott, 2014). The combination of Eq. (6) and Eq. (7) with values $\alpha = Y_i$ and $b = 1$ produces likelihood functions for the i-th observation as follows:

$$\frac{\exp(\mu_i)^{Y_i}}{1+\exp(\mu_i)} \propto \exp(k_i\,\mu_i) \int_0^\infty \left(\exp\frac{-\omega_i\,\mu_i^2}{2}\right) p(\omega_i) d\omega_i \tag{9}$$

By considering the parameter $\boldsymbol{\omega}$ and autocorrelation parameter of SAR ρ, a conditional posterior distribution $\boldsymbol{\beta}$ is obtained, as follows:

$$p(\boldsymbol{\beta}|\rho,\boldsymbol{\omega},\boldsymbol{Y}) \propto p(\boldsymbol{\beta}) \prod_{i=1}^{N} \exp\left(k_i\,\mu_i - \tfrac{\omega_i\,\mu_i^2}{2}\right), \tag{10}$$

$$\propto p(\boldsymbol{\beta}) \exp\left\{-\tfrac{1}{2}(\mathbf{A}\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}(\mathbf{A}\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta})\right\}, \tag{11}$$

with matrix $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, ..., \omega_N)$ of size $(N \times N)$, and vector $\boldsymbol{z} = (k_1/\omega_1, ..., k_N/\omega_N)^{\mathrm{T}}$. By bringing up the prior density function that is normally distributed for parameter $\boldsymbol{\beta}$ with average $\underline{\boldsymbol{\mu}}_\beta$ and variance $\boldsymbol{\Sigma}_\beta$, $p(\boldsymbol{\beta}) \sim \mathcal{N}(\underline{\boldsymbol{\mu}}_\beta, \underline{\boldsymbol{\Sigma}}_\beta)$, the conditional posterior distribution for parameter $\boldsymbol{\beta}$ is Gaussian:

$$p(\boldsymbol{\beta}|\rho,\boldsymbol{\omega},\boldsymbol{Y}) = \mathcal{N}(\overline{\boldsymbol{\mu}}_\beta, \overline{\boldsymbol{\Sigma}}_\beta) \ , \tag{12}$$

with posterior parameter/quantity $\overline{\boldsymbol{\mu}}_\beta = \boldsymbol{\Sigma}_\beta \left(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{A}\boldsymbol{z} + \underline{\boldsymbol{\Sigma}}_\beta^{-1}\underline{\boldsymbol{\mu}}_\beta\right)$ and $\overline{\boldsymbol{\Sigma}}_\beta = \left(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{X} + \underline{\boldsymbol{\Sigma}}_\beta^{-1}\right)^{-1}$. Eq. (12) shows a special advantage of the introduction of the Pólya– Gamma latent variable. By being given $\boldsymbol{\omega}$ conditional posterior distribution $\boldsymbol{\beta}$ is normally distributed. Conditional on $\boldsymbol{\omega}$, where $p(\boldsymbol{\omega})$ is a density function of the random variable $p(\boldsymbol{\omega}) \sim \mathcal{PG}(b, 0)$, $b > 0$ then the identity of the integral applies to all $\in \mathbb{R}$ as in Eq. (7). The conditional posterior distribution for the parameter of SAR ρ is as follows:

$$p(\rho|\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{Y}) \propto |\mathbf{A}| \exp\left\{-\tfrac{1}{2}(\mathbf{A}\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}(\mathbf{A}\boldsymbol{z} - \mathbf{X}\boldsymbol{\beta})\right\} p(\rho) \ , \tag{13}$$

where $p(\rho)$ is a prior density function of SAR parameter ρ. The selection of prior standards for ρ involves a Uniform or beta distribution (LeSage & Pace, 2009). However, the conditional posterior for ρ cannot be reduced to a regular distribution that can be easily sampled. Therefore, the MCMC procedure with the Gibbs sampler algorithm (Ritter & Tanner, 1992) is used in sampling from the conditional posterior for ρ. This can be achieved easily using numerical integration procedures (LeSage & Pace, 2009).

2.4 Distance-Based Spatial Weight Matrix

The spatial weight matrix (**W**) is a matrix of size $(N \times N)$ that illustrates the relationship between locations. The spatial weight matrix **W** is written as follows:

$$\mathbf{W_{ij}} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1N} \\ w_{21} & 0 & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & 0 \end{bmatrix}, \tag{14}$$

and the value of the diagonal matrix **W** is zero (0) because it shows the connection of a location with the location itself. For easy interpretation, the weight value of w_{ij} is the result of row-standardized which is calculated by the following equation (Elhorst, 2014):

$$w_{ij} = \frac{w_{ij}^*}{\sum_{j=1}^{N} w_{ij}^*} \tag{15}$$

The weight value of w_{ij} has values in the range of 0 to 1, which indicates the probability value. The signal strength emitted by a provider tower is spread in all directions. The farther the distance of an area to the center point of the signal transmitter, the more decreased the signal strength received in the area. The spatial weight matrix that can describe the characteristics of signal strength is the distance-based spatial weight matrix. This spatial weight matrix reflects that the spatial effect decreases with the increase in distance caused by increasing geographical resistance (Mills & Fricker, 2011).

The distance-based spatial weight matrix is determined based on the distance between the location coordinate points (longitude and latitude) measured through Euclidean distance between the *i*-th location point and the *j*-th location point which is calculated by the following equation:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \times 111.319 \text{ km} \ , \tag{16}$$

by u_i stating the latitude coordinates at the *i* location and v_i the longitude coordinates at the *i* location. Distance d_{ij} has the unit of degree (°). To get the kilometer (km) unit, it is multiplied by the value of 111.319, because 1° is equal to 111.319km (Irwan & Atmajaya, 2018). There are several types of distance-based spatial weight matrices, namely K-NN, radial distance, power distance, and exponential distance (Zhou & Lin, 2008). This study compares the four (4) weight matrix and then selects one (1) matrix, which is the best use of distance-based spatial weight matrix in internet signal strength modeling.

2.4.1 K-Nearest Neighbor

In the K-NN weight matrix, it must be determined in advance how many K the nearest location is. If the value is $K \in \{1, 2, \ldots, N\text{-}1\}$, then K the closest distance between *i* location and all *j* locations is $N_K \in \{d_{ij[1]}, d_{ij[2]}, \ldots, d_{ij[k]}\}$, with K set of closest j location to i location is $J_K(i) = \{j_{[1]}, j_{[2]}, \ldots, j_{[K]}\}$, so the weight value (w_{ij}^*) is defined as follows:

$$w_{ij}^*(K) = \begin{cases} 1 & \text{if } j \in J_K(i) \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

2.4.2 Radial Distance Weights

In the radial distance matrix, it must be determined in advance what threshold distance (*d*) or bandwidth, which is an important criterion in spatial effect. The distance between the *i*-th location and the *j*-th location that exceeds the threshold distance *d* will not be given the spatial effect weight. The weight value of (w_{ij}^*) in a spatial weight matrix based on radial distance is defined as follows:

$$w_{ij}^* = \begin{cases} 1 & \text{if } 0 \le d_{ij} \le d \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

2.4.3 Power Distance Weights

If there is evidence that the farther the distance, the decrease in spatial effect, the approach to the power distance matrix is by assuming that the weight is a negative power function of distance, defined as follows:

$$w_{ij}^* = d_{ij}^{-r} \ , \tag{19}$$

with *r* is any positive exponent, typically $r = 1$ or $r = 2$.

2.4.4 Exponential Distance Weights

An alternative to the power distance weight matrix is the negative exponential function of distance, defined as follows:

$$w_{ij}^* = \exp(-l \times d_{ij}) \ , \tag{20}$$

with *l* is any positive exponent.

2.5 Distance-Based Spatial Weight Optimization

The value determination of *K, d, r,* and *l* respectively in the K-NN, radial distance, power distance and exponential distance weight matrices are subjective from the researcher's assumption. Therefore, it is not uncommon that the value given does not provide optimal weight. In the study of Jaya, Tantular, & Zulhanif (2017), the number of nearest neighbors of K in the K-NN weight matrix was optimized based on the highest Moran's Index (Moran's I) statistics. However, the value of Moran's I cannot be obtained from the Logistic SAR model that uses the response variable category. In addition to Moran's I, the value that shows the strength of spatial autocorrelation in the Logistic SAR model is the spatial lag coefficient parameter ρ, which has a range of values of -1 to 1. Therefore, by adapting the same concept of the study of Jaya, Tantular, & Zulhanif (2017), the value of *K, d, r,* and *l* in this study is determined based on the highest absolute value of $|\rho|$.

2.6 Spatial Autocorrelation Test for Binary Data

A commonly used statistical test to check whether a variable has the effect of spatial autocorrelation is Moran's I or Geary's contiguity ratio. However, both tests can only be used for continuous data. For category data, especially binary data, testing can be done using the Join Count statistics (Cliff & Ord, 1981). This method measures spatial relationships between similar, same, or different attributes in adjacent locations. Binary variables in this method are expressed into two (2) colors, which are defined as follows:

$$X_i = \begin{cases} 1 & (\boldsymbol{B}lack) \\ 0 & (\boldsymbol{W}hite) \end{cases} \tag{21}$$

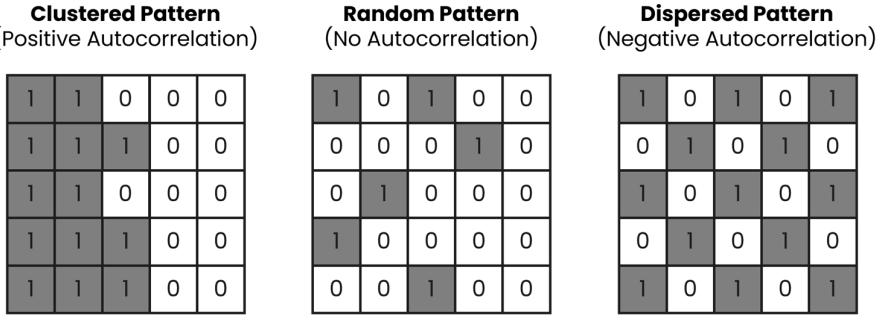


Fig. 3. Indication of Spatial Autocorrelation Based on The Type of Distribution Pattern

Based on Fig. 3, there are three (3) types of joins that may occur, namely BB (Black-Black) joins if both adjacent locations have the black color/value of 1, WW (White-White) joins if both adjacent locations have the white color/value of 0. These two joins show a positive autocorrelation. Meanwhile, BW (Black-White) joins occur if the two adjacent locations have different colors or are valued at 0 and 1. The third join shows negative autocorrelation. The Join Count statistic calculates the sum of the three observed combines, obtained by the following equation:

$$BB = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} X_i X_j \ , \tag{22}$$

$$WW = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (1 - X_i)\left(1 - X_j\right) \ , \tag{23}$$

$$BW = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \left(X_i - X_j\right)^2 \ , \tag{24}$$

where w_{ij} is an element on the spatial weight matrix that has been row-standardized. For hypothesis testing on this method, a normal approach with a *Z*-statistic test for each join can be used.

Hypothesis:

$H_0 : \rho = 0$ (there is no spatial autocorrelation)

$H_1 : \rho \neq 0$ (there is spatial autocorrelation)

Z-statistic test:

$$Z = \frac{Observed - Expected}{Standard\ Deviation\ of\ Expected} \tag{22}$$

with

Expected:	Standard Deviation of Expected (Standard Error):
$E(BB) = k \cdot p_B^2$	$\sqrt{Var(BB)} = \sqrt{(k \cdot p_B^2) + (2m \cdot p_B^3) - (k + 2m)p_B^4}$
$E(WW) = k \cdot p_W^2$	$\sqrt{Var(WW)} = \sqrt{(k \cdot p_W^2) + (2m \cdot p_W^3) - (k + 2m)p_W^4}$
$E(BW) = 2k \cdot p_B \cdot p_W$	$\sqrt{Var(BW)} = \sqrt{2(k + m)p_B^2 \cdot p_W^2 - 4(k + 2m)p_B^2 \cdot p_W^2}$
	and $m = \frac{1}{2}\sum_{i=1}^{N} k_i(k_i - 1)$

where k is the number of *joins*, p_B is the expected proportion of black attribute, and p_W is the expected proportion of the white attribute. The test criterion is to reject the null hypothesis if the value is $|Z| \geq Z_{1-(\alpha/2)}$ or p-value ≤ significance level (α). Do not reject the null hypothesis in any other conditions.

2.7 Model Evaluation

A tool that can be used to evaluate the performance of a classification model is the confusion matrix (Stehman, 1997). Here is the table from the confusion matrix for two (2) classes/categories:

Table 2
Confusion Matrix

Predicted Class	**Actual Class**	
	Positive (1)	**Positive (1)**
Positive (1)	True Positive (TP)	False Positive (FP), Type I Error
Negative (0)	False Negative (FN), Type II Error	True Negative (TN)

with

TP, TN : the number of observations whose actual classes are the same as the prediction classes.

FP, FN : the number of observations whose actual classes are not the same as the prediction classes.

The size of the model performance produced from the confusion matrix is accuracy, precision, recall, and F1-score which are respectively calculated by the following equation:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad , \tag{23}$$

$$precision = \frac{TP}{TP+FP} \quad , \tag{24}$$

$$recall = \frac{TP}{TP+FN} \quad , \tag{25}$$

$$F1-score = 2 \times \frac{precision \times recall}{(precision+recall)} \quad , \tag{26}$$

here the four performance sizes of the model have values in the range of 0 to 1. The higher the performance size of the model, the better the model produced.

2.8 Steps of Data Analysis

Data processing is carried out using R software version 4.3.0. The data analysis steps in this study were:

1. Data exploration based on distribution maps for all villages in Tasikmalaya Regency.
2. Calculate the distance matrix between village office locations with Eq. (16).
3. Create a grid of values of *K, d, r,* and *l* respectively for the optimization process of K-NN, radial distance, power distance, and exponential distance weight matrices.
4. Form a spatial weight matrix as in Eq. (14) and (15), with the K-NN, radial distance, power distance, and exponential distance methods as in Eqs. (17-20).
5. Carried out spatial autocorrelation testing with the Join Count statistical test as in Eq. (25).
6. Logistic SAR model estimation using the Bayesian MCMC approach:
   a. Forms likelihood functions as in Eq. (9).
   b. Determining the prior, this study used the prior density function from the normal distribution (Gaussian) for parameter $\boldsymbol{\beta}$ and the prior density function involving uniform/ beta distribution for parameter ρ.
   c. Update the value of $\boldsymbol{\omega}$ from the conditional posterior distribution function $p(\boldsymbol{\omega} \mid \boldsymbol{\beta}, \rho, \boldsymbol{Y})$ as in Eq. (8).
   d. Update the value of $\boldsymbol{\beta}$ from the conditional posterior distribution function $p(\boldsymbol{\beta} \mid \rho, \boldsymbol{\omega}, \boldsymbol{Y})$ as in Eq. (12).
   e. Update the value of ρ using the Gibbs step from the conditional posterior distribution $p(\rho \mid \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{Y})$ as in Eq. (13).
   f. The process of estimating the parameter value of $\boldsymbol{\omega}$, $\boldsymbol{\beta}$, and ρ carried out iteratively with the MCMC sampling algorithm (Gibbs Sampler) procedure.
7. Optimization of values of *K, d, r,* and *l* respectively for K-NN, radial distance, power distance, and exponential distance weight matrices based on the highest absolute value of $|\rho|$.

8. Selection of the best distance-based weight matrix based on the evaluation of model performance through the highest value of accuracy, precision, recall, and F1-score as in Eqs. (26-29).
9. Analysis and interpretation of the best Logistic SAR Model.

3. Result and Discussion

3.1 Descriptive Analysis

Based on the proportion of the number of villages according to signal strength in Tasikmalaya Regency in Figure 1, the following is a detailed map of the signal strength distribution:

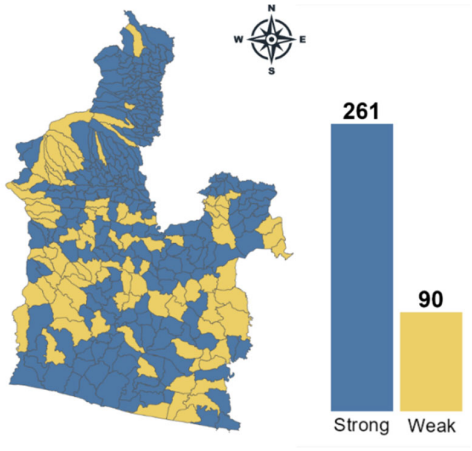
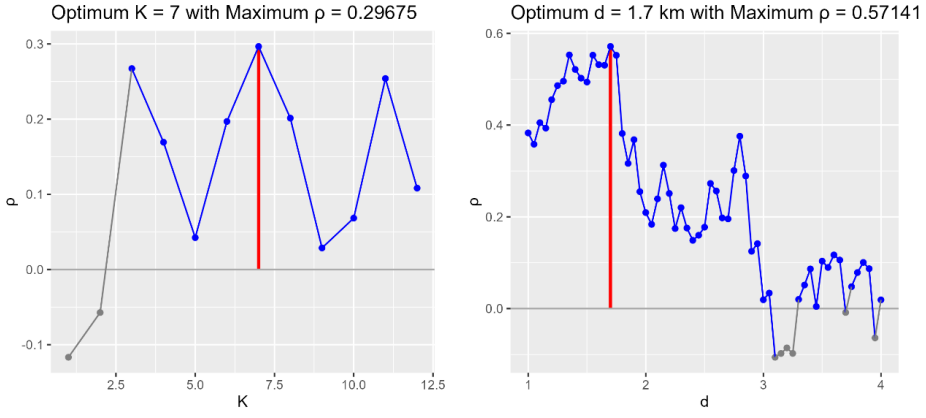


Fig. 4. Map of Cellular Phone Signal Strength Distribution in Villages in Tasikmalaya Regency in 2022

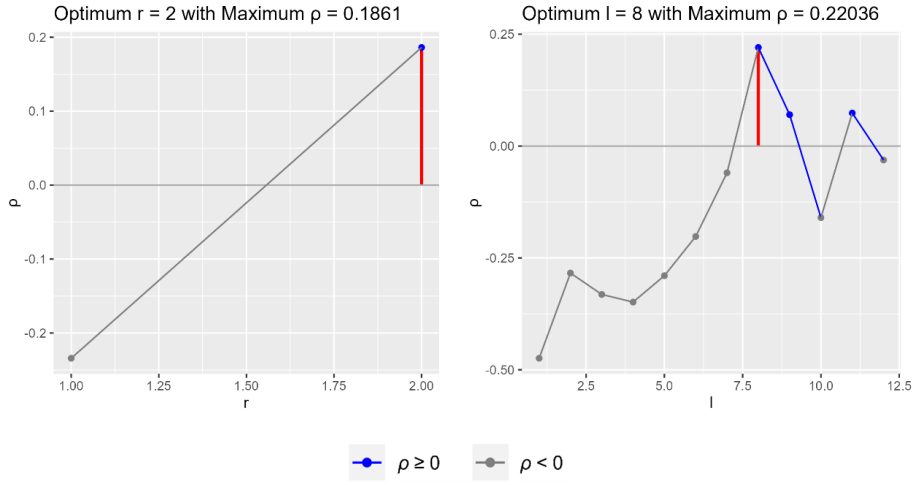
Based on Fig. 4, the 90 villages that have weak signals in Tasikmalaya Regency tend to be spread in the southern region. These villages tend to gather into several groups. Descriptively, this shows that the signal strength in Tasikmalaya Regency is indicated by a tendency for positive spatial effects.

3.2 Logistic Spatial Autoregressive Model

The first stage of Logistic SAR modeling is to form a spatial weight matrix. In this study, four (4) distance-based spatial weight matrices were used, namely K-NN, radial distance, power distance, and exponential distance weight matrices. Based on the characteristics of the signal strength as well as the previous descriptive data analysis, it showed the presence of positive spatial autocorrelation. Therefore, the process of weight optimizing for the four (4) distance-based spatial weight matrices was based on the positive value of the highest estimated parameter ρ. The distance-based weight optimization process in this study uses the grid of value of $K = 1, 2, \ldots, 12$ nearest neighbors for the K-NN weight matrix, the grid of value of $d = 1, 1.05, 1.1, \ldots, 4$km for the radial distance weight matrix, and the grid of value of $r = 1, 2$ for the power distance weight matrix, and the grid of value of $l = 1, 2, \ldots, 12$ for the exponential distance weight matrix. The results of distance-based weight optimization can be seen in Fig. 5:



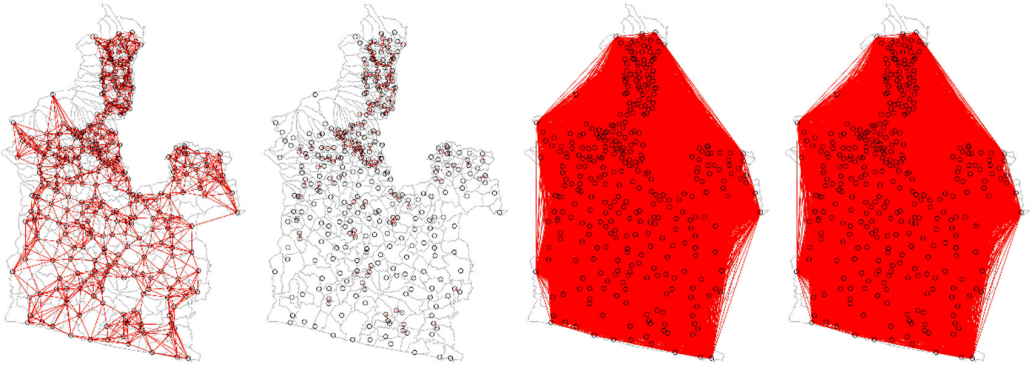
(a) (b)



(c) (d)

Fig. 5. Plot Value of ρ with Value of K (a), d (b), r (c), and l (d)

Based on the plot in Fig. 5, the most optimal weight of the K-NN weight matrix is found in the value of $K = 7$ nearest neighbor villages for internet signal strength in Tasikmalaya Regency, with the highest value of ρ of 0.29675. For the radial distance weight matrix, the most optimal weight is found at the threshold distance $d = 1.7$km, with the highest value of ρ of 0.57141. Next, the most optimal weight of the power distance weight matrix is found at the value of $r = 2$, with the highest value of ρ of 0.18610. The most optimal weight of the exponential distance weight matrix is found at the value of $l = 8$, with the highest value of ρ of 0.22036. The optimization results of the four (4) distance-based spatial weight matrices are visualized in the form of maps in Fig. 6.



(a) (b) (c) (d)

Fig. 6. Visualization of K-NN (a), Radial Distance (b), Power Distance (c), and Exponential Distance (d) Optimum Weight Matrices

The spatial weight matrix in Figure 6 is a (351 × 351) sized matrix which was formed based on 351 location points of village offices in Tasikmalaya Regency. For the K-NN and radial distance weight matrices, the location points that are connected to other location points with a red line are given a weight value of one (1), and those that are not connected by the red line are given a weight value of zero (0). Whereas in the power distance and exponential distance weight matrix, all location points are connected to each other with a red line because all locations have a weight value of more than zero ($w_{ij} > 0$).

Furthermore, spatial autocorrelation testing was carried out on the Y response variable for the four (4) most optimal distance-based weight matrices. The test results can be seen in Tables 3, 4, 5, and 6:

Table 3
Joint Count Test Results Using 7-Nearest Neighbor Weight Matrix

Join	Join Count	Expected	Std. Deviation	Z-value	Pr (>|Z|)
Strong-Strong	105.8571	96.9429	1.2886	6.9178	2.2935e-12*
Weak-Weak	19.1429	11.4429	0.9540	8.0716	3.4689e-16*
Strong-Weak	50.5000	67.1143	1.8999	-8.7450	1.1148e-18*
Total	**175.5000**				

*) Significant at $\alpha = 0.05$

Table 4
Joint Count Test Results Using Radial Distance Weight Matrix with $d = 1.7$km

Join	Join Count	Expected	Std. Deviation	Z-value	Pr (>|Z|)
Strong-Strong	85.2750	152.8378	0.4200	-160.8465	0*
Weak-Weak	8.7500	18.0405	2.0502	-4.5315	2.9278e-06*
Strong-Weak	17.4750	105.8108	4.6726	-18.9052	5.1672e-80*
Total	**111.5000**				

*) Significant at $\alpha = 0.05$

Table 5
Joint Count Test Results Using Power Distance Weight Matrix with $r = 2$

Join	Join Count	Expected	Std. Deviation	Z-value	Pr (>|Z|)
Strong-Strong	106.5013	96.9429	1.0459	9.1388	3.1575e-20*
Weak-Weak	15.6816	11.4429	0.7559	5.6073	1.0274e-08*
Strong-Weak	53.3170	67.1143	1.5048	-9.1689	2.3885e-20*
Total	**175.5000**				

*) Significant at $\alpha = 0.05$

Table 6
Joint Count Test Results Using Exponential Distance Weight Matrix with $l = 8$

Join	Join Count	Expected	Std. Deviation	Z-value	Pr (>|Z|)
Strong-Strong	100.1815	96.9429	0.5600	5.7829	3.6705e-09*
Weak-Weak	13.0525	11.4429	0.2243	7.1766	3.5737e-13*
Strong-Weak	62.2660	67.1143	0.4364	-11.1097	5.6279e-29*
Total	**175.5000**				

*) Significant at $\alpha = 0.05$

The results of the Join Count test above were seen based on the *p-value* in the join categories of strong-strong and weak-weak. The reason is that the optimization results in the four (4) distance-based spatial weight matrices (Figure 5) produce the value of $\rho > 0$ which indicates a positive spatial autocorrelation. Based on Tables 3, 4, 5, and 6, the variable of cellular phone/ handphone signal strength in the village in Tasikmalaya Regency shows a significant positive spatial autocorrelation. This is because the join categories of strong-strong and weak-weak for each distance-based spatial weight matrix produce a *p-value* of less than 5%. Therefore, the Logistic SAR modeling can be performed. The results of estimating the coefficient parameters from the Logistic SAR model with the Bayesian MCMC approach for the four (4) distance-based spatial weight matrices can be seen in Table 7 below:

Table 7
The Results of Estimating Parameter of the Logistic SAR Model Based on K-NN, Radial Distance, Power Distance, and Exponential Distance Weight Matrix

Coefficient	Distance-Based Spatial Weight Matrix			
	7-NN	Radial Distance ($d = 1.7$)	Power Distance ($r = 2$)	Exponential Distance ($l = 8$)
(Intercept)	1.33058	1.01106	1.50062	1.47851
Distance of the nearest provider tower to the village office/center of the village/village settlement (X_1)	-0.00034	-0.00025	-0.00038	-0.00039
Height difference between the nearest provider tower placement and village office location (X_2)	-0.00628	-0.00581	-0.00642	-0.00701
Rho (ρ)	0.29675	0.57141	0.18610	0.22036

Based on the results of estimating the coefficient parameter of the Logistic SAR model in Table 7, predictions are made and then a confusion matrix is formed based on the comparison between actual categories/classes with prediction results. Confusion matrix for the four (4) distance-based spatial weight matrices that can be seen in Tables 8, 9, 10, and 11:

Table 8
Confusion Matrix of Logistic SAR Model Using 7-Nearest Neighbor Weight Matrix

Predicted Class	Actual Class		Total
	Strong (1)	Weak (0)	
Strong (1)	245	72	**317**
Weak (0)	16	18	**34**
Total	**261**	**90**	**351**

Table 9
Confusion Matrix of Logistic SAR Model Using Radial Distance Weight Matrix with $d = 1.7$km

Predicted Class	Actual Class		Total
	Strong (1)	Weak (0)	
Strong (1)	242	67	**309**
Weak (0)	19	23	**42**
Total	**261**	**90**	**351**

Table 10
Confusion Matrix of Logistic SAR Model Using Power Distance Weight Matrix with $r = 2$

Predicted Class	Actual Class		Total
	Strong (1)	Weak (0)	
Strong (1)	245	75	**320**
Weak (0)	16	15	**31**
Total	**261**	**90**	**351**

Table 11
Confusion Matrix of Logistic SAR Model Using Exponential Distance Weight Matrix with $l = 8$

Predicted Class	Actual Class		Total
	Strong (1)	Weak (0)	
Strong (1)	245	75	**320**
Weak (0)	16	15	**31**
Total	**261**	**90**	**351**

To compare the performance of the Logistic SAR model based on the use of the four (4) distance-based weight matrices, the accuracy, precision, recall, and F1-score values of each category are calculated, in percentage (%), which are presented in Table 12:

Table 12
Logistic SAR Model Performance Level Based on K-NN, Radial Distance, Power Distance, and Exponential Distance Weight Matrices

Positive Class	Performance Value	Distance-Based Spatial Weight Matrix				Number of Observations
		7-NN	Radial Distance ($d = 1.7$)	Power Distance ($r = 2$)	Exponential Distance ($l = 8$)	
Strong (1)	**Accuracy**	74.93%	75.50%	74.07%	74.07%	**261**
	Precision	77.29%	78.32%	76.56%	76.56%	
	Recall	93.87%	92.72%	93.87%	93.87%	
	F1-Score	84.78%	84.91%	84.34%	84.34%	
Weak (0)	**Accuracy**	74.93%	75.50%	74.07%	74.07%	**90**
	Precision	52.94%	54.76%	48.39%	48.39%	
	Recall	20.00%	25.56%	16.67%	16.67%	
	F1-Score	29.03%	34.85%	24.79%	24.79%	
Average F1-Score		**56.90%**	**59.88%**	**54.57%**	**54.57%**	**351**

*Information:
Highest value
Lowest value

Based on Table 12, modeling the internet signal strength category in Tasikmalaya Regency, spatially, is more precise and relevant using the radial distance weight matrix. This is shown by the accuracy, precision, recall and F1-score values in the radial distance weight matrix have the highest level of performance compared to the other three (3) distance-based spatial weight matrices, except for recall, when the strong positive class has the lowest value. In addition, the average F1-score on the radial distance weight matrix has the highest value, which is 59.88%. This means that overall, the Logistic SAR model using this spatial weight matrix is the best model. On the other hand, the weight matrices of power distance and exponential distance have the exact same level of performance. This may be because the weighting method of the two (2) matrices tends to be the same, only different in the function.

Logistic SAR modeling of internet signal strength category in Tasikmalaya Regency in the four (4) distance-based spatial weight matrices produce a high level of performance in predicting the category of strong signals, but low in predicting the category of weak signal. This may be because the amount of observation between categories has a fairly high difference; with a respective proportion of 74.36% of the villages with strong signals and 25.64% of the villages with weak signals.

3.3 Analysis and Interpretation of The Best Logistic Spatial Autoregressive Model

Based on the selection results of the best distance-based spatial weight matrix, that is the radial distance weight matrix with a threshold distance of d = 1.7km, the Logistic SAR model of cellular phone/handphone signal strength based on the factor of distance and height difference between the placement of the nearest provider tower and the location of the village office in Tasikmalaya Regency with the Bayesian MCMC approach, can be written into the following equation:

$$\hat{Y}_i^* = logit\big(\pi(X_i)\big) = 0{,}57141 \sum_{j=1, i\neq j}^{N} w_{ij} Y_j^* + 1.01106 - 0.00025\, X_{i1} - 0.00581\, X_{i2} \quad (27)$$

Based on Eq. (30), the coefficient of lag spatial (ρ) has a value of 0.57141, which shows that spatial has a positive effect on the response variable of cellular phone/handphone signal strength. This means that the internet signal strength of villages in Tasikmalaya Regency is affected by surrounding villages that are 1.7km or less, amounting to 57.141%. Meanwhile, the factor of distance (X_1) and height (X_2) difference between the location of the nearest provider tower and the location of the nearest village office negatively affects the Y response variable. This means that the closer the distance (both horizontally and vertically) to the location of the provider tower with the location of the village office, the more likely the village is to have a strong signal. To look closely, the height difference between the location of the provider tower with the location of the village office has a greater effect than the horizontal distance.

4. Conclusions

The use of a weight matrix with radial distance method in Logistic SAR modeling of internet signal strength in Tasikmalaya Regency is the most suitable. This is based on the confusion matrix generated by the radial distance weight matrix that has the highest average level of performance compared to the other distance-based spatial weight matrices. That is, the radial distance weight matrix is more relevant and can best describe the characteristics of signal strength that are radius. The result of the analysis showed that the strength of positive spatial autocorrelation on the internet signal strength of villages in Tasikmalaya Regency was 57.141%, with an optimum threshold distance of d = 1.7km. In addition, the factor of distance and height difference between the placement of the nearest provider tower and the village office equally have negative effects. However, the height difference between the location of the provider tower with the location of the village office has a greater effect than the horizontal distance. The model is expected to be useful as a reference for government agencies in Indonesia, especially the Tasikmalaya Regency Government, in developing strategies for locating signal transmission towers in rural areas. To ensure that the tower's signal reaches the villages equitably, the concerned government can prioritize placing the signal transmitting tower near the village office/center of the village/village settlement while focusing on elevation, rather than just horizontal distance. Aside from that, villages in Tasikmalaya Regency tend to have similar signal strengths to neighboring villages that are less than 1.7km distant.

Another finding, the evaluation results of Logistic SAR modeling of the internet signal strength category in Tasikmalaya Regency with the Bayesian MCMC approach have a high-performance level in predicting the category of strong signals, but a low-performance level in predicting the category of weak signals. This may be because the amount of observation between categories has a fairly high difference; with a respective proportion of 74.36% of the villages with strong signals and 25.64% of the villages with weak signals.

Acknowledgement

The authors would like to thank the Rector of Universitas Padjadjaran, who offered financial help to disseminate studies reports. This research is part of the RPLK scheme with the contract No: 1549/UN6.3.1/PT.00/2023 and Academic Leadership Grant (ALG) contract No: 1549/UN6.3.1/PT.00/2023. We are also thankful for the discussion on social media analytics through the RISE_SMA project funded through the European Union year 2019-2024 and also the study center of modeling and computation Universitas Padjadjaran.

References

Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes, Models and Applications.* London: Pion.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels.* Berlin: Springer.

Indonesia Internet Service Provider Association. (2022). *Survei Profil Internet Indonesia [Indonesian Internet Profile Survey].* Jakarta: Asosiasi Penyelenggara Jasa Internet Indonesia. Retrieved February 9, 2023, from https://apjii.or.id/download/223be323e6e3778c46872e727a936ea4

Irwan, & Atmajaya, D. (2018). Sistem Informasi Pencarian Lokasi Perguruan Tinggi di Makassar [Information System for University Location Search in Makassar]. *ILKOM Jurnal Ilmiah, X*(2), 232-236. doi:10.33096/ilkom.v10i2.251.232-236

Iskandar, A. H. (2020). *SDGs Desa Percepatan Pencapaian Tujuan Pembangunan Nasional Berkelanjutan [SDGs Village Acceleration of Achievement of Sustainable National Development Objectives]*. Jakarta: Yayasan Pustaka Obor Indonesia.

Jaya, I. G., Tantular, B., & Zulhanif. (2017, March 18). Optimalisasi Matriks Bobot Spasial Berdasarkan K-Nearest Neighbor dalam Spasial Lag Model [Optimization of Spatial Weight Matrix Based on K-Nearest Neighbor in Spatial Lag Model]. *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya II (KNPMP II)*, 104-111.

Krisztin, T., & Piribauer, P. (2020). A Bayesian Spatial Autoregressive Logit Model with an Empirical Application to European Regional FDI Flows. *Empirical Economics, LXI*, 231–257. doi:10.1007/s00181-020-01856-w

Lee, W. C.-Y. (2006). *Mobile Cellular Communication: Analog and Digital Systems* (3rd ed.). New York: McGraw-Hill.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics.* New York: Taylor & Francis Group, LLC.

Mills, J. B., & Fricker, J. D. (2011). Integrated Analysis of Economic Impacts of Bypasses on Communities: Panel Data Analysis and Case Study Interviews. *Transportation Research Record: Journal of the Transportation Research Board, MMCCXLII*(1), 114-121. doi:10.3141/2242-14

Ministry of Home Affairs Republic of Indonesia. (2022). Keputusan Menteri Dalam Negeri Nomor 050-145 Tahun 2022 Tentang Pemberian dan Pemutakhiran Kode, Data Wilayah Administrasi Pemerintahan, dan Pulau Tahun 2021. Jakarta: Ministry of Home Affairs Republic of Indonesia.

Ministry of Villages, Development of Disadvantaged Regions, and Transmigration Republic of Indonesia. (2016). Peraturan Menteri Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Republik Indonesia Nomor 2 Tahun 2016 Tentang Indeks Desa Membangun. Jakarta: Ministry of Villages, Development of Disadvantaged Regions, and Transmigration Republic of Indonesia.

Ministry of Villages, Development of Disadvantaged Regions, and Transmigration Republic of Indonesia. (2022). Keputusan Menteri Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Republik Indonesia Nomor 80 Tahun 2022 Tentang Status Kemajuan dan Kemandirian Desa Tahun 2022. Jakarta: Ministry of Villages, Development of Disadvantaged Regions, and Transmigration Republic of Indonesia.

Pinkse, J., & Slade, M. E. (1998). Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models. *Journal of Econometrics, LXXXV*(1), 125-154. doi:10.1016/S0304-4076(97)00097-3

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. *Journal of the American Statistical Association, CVIII*(504), 1339-1349. doi:10.1080/01621459.2013.829001

Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Association, LXXXVII*(419), 861-868. doi:10.2307/2290225

Stehman, S. V. (1997). Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sensing of Environment, LXII*(1), 77–89. doi:10.1016/S0034-4257(97)00083-7

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography, XLVI*, 234-240. doi:10.2307/143141

Windle, J., Polson, N. G., & Scott, J. G. (2014). Sampling Pólya–Gamma Random Variates: Alternate and Approximate Techniques. *arXiv:1405.0506*. doi:10.48550/arXiv.1405.0506

Zhou, X., & Lin, H. (2008). Spatial Weights Matrix. *Encyclopedia of GIS*. doi:10.1007/978-0-387-35973-1_1307

© 2024 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).