Application of the Geographically Weighted Negative Binomial Regression Method to Tuberculosis Cases in North Sumatra Province in 2024

N L Meilawati, N H Nufus, R Rahma, T J Tinambunan, F Wicaksono*

Politeknik Statistika STIS, Jakarta, Indonesia

***Corresponding author's email: febri@stis.ac.id

Abstract. Tuberculosis (TB) is a chronic, contagious disease caused by Mycobacterium tuberculosis and one of the leading causes of death worldwide, with approximately 1.2 million deaths occurring annually. According to the World Health Organization, Indonesia is the second-largest tuberculosis country after India. In 2024, North Sumatra, a province in Indonesia, had the highest number of TB cases on Sumatra Island, surpassing the national average and ranking third in the country. The number of TB cases in North Sumatra is census data and overdispersed, with spatial influences. This study examines the characteristics, spatial distribution, and determinants of tuberculosis cases across districts and cities in North Sumatra in 2024. Therefore, the method used is Geographically Weighted Negative Binomial Regression, which produces local parameters and can be used to examine the influence of independent variables on tuberculosis cases spatially. Environmental factors can influence disease-causing agents and have a significant role in the transmission of infectious diseases, including Tuberculosis. This study includes rainfall to represent Environmental variables. Research related to Tuberculosis using environmental variables in North Sumatra has never been conducted before. The results show that GWNBR forms eight regional groups based on significant variables. Rainfall and per capita expenditure variables significantly influence all districts, the percentage of BCG immunizations, the percentage of smoking population, and health fund allocation significantly influence several districts.
Keywords: GWNBR, Negative Binomial, Overdispersion, Poisson, Tuberculosis,

1. Introduction

Sustainable Development Goal 3.3 establishes the End Tuberculosis Strategy, with a target of reducing the number of deaths by 90% and reducing Tuberculosis cases by 80% by 2030 [1]. Tuberculosis is one of the leading causes of death worldwide. In some countries, treating tuberculosis remains a challenge and remains a global health emergency due to its impact on high morbidity and mortality rates. Tuberculosis ranks second as the most infectious disease in children and adults after HIV. Approximately 1.2 million deaths occur annually from tuberculosis [2].

Tuberculosis (TB) is a chronic, contagious disease caused by Mycobacterium tuberculosis. This disease can be easily transmitted under certain conditions, such as when the immune system is weakened. Furthermore, tuberculosis generally affects lower socioeconomic groups. People with tuberculosis typically experience several symptoms related to the respiratory system, ranging from coughing, coughing up blood, pain when breathing, to chest pain. However, this disease can also occur without any symptoms at all [3].

According to the World Health Organization (WHO), Indonesia is the country with the second-highest tuberculosis caseload after India, accounting for 10% of the total, followed by China, the Philippines, Pakistan, Nigeria, Bangladesh, and the Democratic Republic of the Congo, respectively [4]. According to the Global TB Report 2024, Indonesia ranks second in the world in terms of TB caseload after India. There are an estimated 1,090,000 TB cases and 125,000 deaths annually, which equate to approximately 14 deaths every hour. These statistics underscore the urgency of increasing prevention and treatment efforts throughout Indonesia [5].

According to 2024 Ministry of Health data, North Sumatra Province recorded 51,827 TB cases, making it the province with the highest number of cases on Sumatra Island and ranking third nationally. North Sumatra Province has striking social, cultural, and economic differences between its regions. This condition allows for differences in characteristics that influence the number of cases within the province [6].

To understand these differences in characteristics, it is necessary to review the factors that influence the spread of tuberculosis. According to Blum (1974), tuberculosis is influenced by environmental factors, health services, lifestyle, and heredity [7]. First, high rainfall causes the physical environment of the house to be humid and the temperature to decrease. This increases the growth of Mycobacterium bacteria and increases the transmission of tuberculosis [8]. Second, high population density will accelerate the increase in tuberculosis because tuberculosis can be transmitted through the air [9]. Third, smoking is a significant cause of lung disease and reduces the body's immune system, thus increasing the risk of tuberculosis transmission [10]. Fourth, adequate allocation of health funds supports various aspects of prevention, such as early detection, treatment, training of health workers, and the provision of drugs [11]. Fifth, per capita expenditure can increase an individual's risk of developing tuberculosis [12]. Sixth, a high average length of schooling influences the decline in tuberculosis [13]. Seventh, adequate health facilities can support the reduction of tuberculosis cases [14]. Finally, early BCG immunization will reduce tuberculosis [15]. Given the numerous causal factors and high TB cases, an appropriate analysis method is required to achieve more accurate results. Data on the number of TB cases in North Sumatra is a type of count data that generally follows a Poisson distribution, but often experiences overdispersion. To address this issue, negative binomial regression is used because it is more flexible in handling overdispersion or significant data variations. Spatial aspects also influence tuberculosis incidence. Therefore, the GWNBR model is a reasonably effective method for estimating count data using a negative binomial distribution that exhibits spatial heterogeneity.

Several previous studies have also used the GWNBR method to analyze TB cases. For example, research by Mumtaz and Utomo [16]. However, both studies focused on tuberculosis cases on the island of Java and did not include physical environmental variables such as rainfall. Unlike previous studies, this study focuses on the number of tuberculosis cases in districts/cities in North Sumatra in 2024. This study also includes a physical environmental variable, namely rainfall. Environmental factors can influence disease-causing agents and play a significant role in the transmission of infectious diseases, including Tuberculosis [17]. Research related to Tuberculosis using environmental variables in North Sumatra has never been conducted before.

Based on these problems, this study aims to identify the characteristics and distribution patterns of Tuberculosis cases in North Sumatra in 2024 and analyze the spatial influence, as well as the variables that influence the number of tuberculosis cases in each district/city in North Sumatra Province in 2024. Regional mapping will be carried out based on significant variables obtained through the results of the analysis using the GWNBR method. This study is expected to contribute to increasing public knowledge regarding tuberculosis. It can be a basis for policy formulation by policymakers so that prevention efforts can be carried out more effectively and on target, based on influential factors in each region, and support the achievement of the Sustainable Development Goals (SDGs) 2030 targets.

2. Methodology

The data used in this study is secondary data with units of observation consisting of 33 regencies/municipalities in North Sumatra. Based on the literature review and available data sources, the following are the variables used in this study.

Table 1. List of Variables Used.

Notation	Variable	Data Sources
Y	Tuberculosis Cases	BPS North Sumatra’s website
X1	Rain Intensity	Google earth Engine
X2	Population Density	BPS North Sumatra’s website
X3	Government Health Spending	North Sumatra’s Government Website
X4	Per Capita Expenditure	BPS-Statistics Indonesia’s website
X5	Average Years of Schooling	BPS North Sumatra’s website
X6	Percentage of Smokers	Statistik Kesejahteraan Rakyat Provinsi Sumatra Utara 2024
X7	Health Facilities	DJSN’s official website
X8	Percentage of BCG	Sumatra Utara dalam Angka 2025

The descriptive statistics will be presented using charts and tables to explore the data, as well as thematic maps to illustrate the distribution of Tuberculosis and its influencing variables. Inferential analysis is conducted using modeling techniques that explain the effect of independent variables on the dependent variable. The analytical steps are as follows:

1. Describing the variables suspected to influence the number of Tuberculosis cases in North Sumatra in 2024.
2. Checking for multicollinearity.
3. Testing the equidispersion assumption.
4. Testing for spatial effects, including both autocorrelation and heterogeneity.
5. Determining the bandwidth and spatial weight matrix.
6. Building the Geographically Weighted Negative Binomial Regression (GWNBR) model.
7. Testing the overall significance of parameters (simultaneous test).
8. Testing the individual significance of parameters (partial test).
9. Classifying regencies/municipalities based on significant variables.

2.1. Tuberculosis

Tuberculosis is a chronic and contagious disease caused by Mycobacterium tuberculosis. It can be easily transmitted under certain conditions, such as when the immune system is weakened. Additionally, tuberculosis commonly affects individuals from lower socioeconomic groups. People with tuberculosis typically experience symptoms related to the respiratory system, including coughing, coughing up blood, chest pain, and breathing discomfort. However, the disease can also occur without any noticeable symptoms.

2.2. Overdispersion

Overdispersion is a condition in which the variance is greater than the mean in a Poisson regression model. Overdispersion can be identified by calculating the dispersion parameter. A dispersion parameter equal to 1 indicates equidispersion, less than 1 indicates underdispersion, and greater than 1 indicates overdispersion. Dispersion can be tested by dividing the deviance and the Pearson Chi-square by their respective degrees of freedom.

2.3. Spatial Autocorrelation

Spatial autocorrelation describes how the value of an observation in a given area depends on the values of observations in neighboring areas. The relationship between these observations can be analyzed using spatial autocorrelation and is commonly measured with Moran's Index. The formula for Moran's Index used is as follows:

$$I = \frac{n\sum_i \sum_j W_{ij}(x_i-\bar{x})(x_j-\bar{x})}{W\sum(x_i-\bar{x})^2} \quad (1)$$

where:

n : number of observations

x_i : the observed value in the *i*-th region

x_j : the observed value in the *j*-th region

$\underline{x}$: the average value in region n

W_{ij} : the weight matrix between region *i* and region *j*

W : the total sum of all values in the correlation matrix

2.4. Spatial Heterogeneity

Spatial heterogeneity is a common issue in spatial structures, caused by differences or instability in the independent variables across each spatial unit of analysis. Spatial heterogeneity is also associated with general heterogeneity that arises when there are omitted variables or other specification errors. This study applies a test for spatial heterogeneity using the Breusch-Pagan test. The formula used is as follows:

$$BP = \frac{1}{2} f^T \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T f \sim X^2_{(k-1)} \quad (2)$$

with $\boldsymbol{f} = (\boldsymbol{f_1}, \boldsymbol{f_2}, \ldots, \boldsymbol{f_n})^{\boldsymbol{T}}$ obtained by $f_i = (\frac{e_i^2}{\sigma^2} - 1)$ and $e_i = y_i - \hat{y}_i$

where

e_i^2 : the squared error for the *i*-th observation

Z : the standardized vector matrix for each observation region, sized n × *p*

$\hat{\sigma}^2$**:** the variance of the dependent variable

2.5. Optimum Bandwidth

The selection of the optimal bandwidth affects the resulting parameter coefficients. Bandwidth acts as a "smoothing parameter" — the higher the bandwidth value, the better the model fit tends to be. However, an oversmoothed model may produce homogeneous parameter values, while an undersmoothed model may lead to heterogeneous parameter values. The optimal bandwidth can be selected by considering the smallest CV (cross-validation) value or the lowest AIC (Akaike Information Criterion).

2.6. Geographically Weighted Negative Binomial Regression (GWNBR)

GWNBR is a statistical method used to model count data that exhibits overdispersion and contains spatial heterogeneity. The GWNBR model produces local parameters for each observation area, with parameter values that may differ across regions. The formula for the GWNBR model is as follows [18]:

$$E[y_i] = \hat{\mu}_i = exp\,\{\,\beta_0(u_i, v_j) + \sum_{i=1}^{p} \beta_k(u_i, v_j)x_{ik} + \theta(u_i, v_j)\} \quad (3)$$

where :
y_i: the value of the *i*-th observation
x_{ik} : the value of the *k*-th independent variable in the observation region (u_i, v_j)
(u_i, v_j) : the location point of the *i*-th observation
$\beta_k\,(u_i, v_j)$: the regression coefficient of the *k*-th independent variable for each region (u_i, v_j)
$\theta\,(u_i, v_j)$: the dispersion parameter for each region $(\boldsymbol{u_i}, \boldsymbol{v_j})$]

3. Results and Discussion

3.1. Descriptive Analysis

North Sumatra Province is located between 1° and 4° North Latitude and 98° and 100° East Longitude, with an area of approximately 71,680 km2. Administratively, the province is divided into 33 districts/cities, with eight cities and 25 regencies. The territorial division of North Sumatra Province can be described as follows:

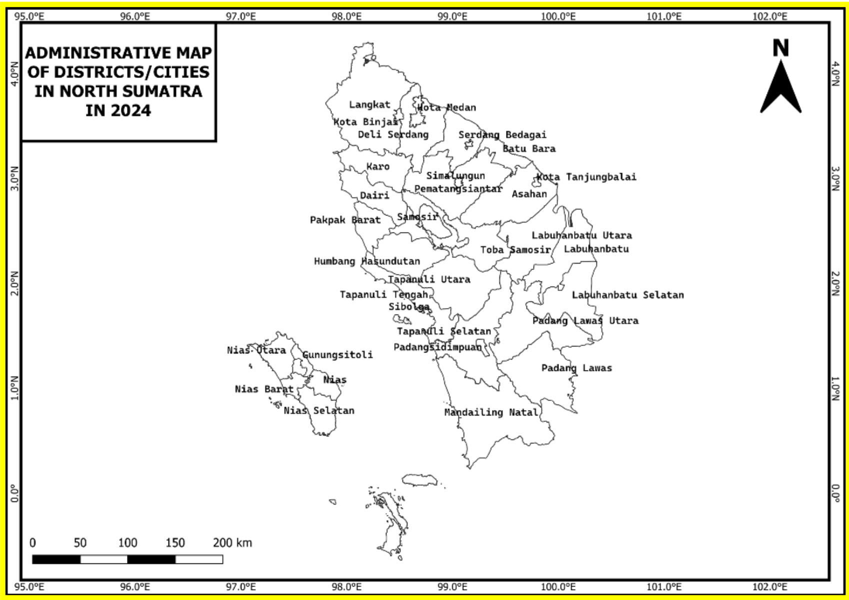

Figure 2. Administrative Map of Districts/Cities in North Sumatra Province in 2024.

Before conducting further analysis, the researcher first examined the data condition through descriptive analysis using a statistical summary table and data distribution using a thematic map in North Sumatra Province. The statistical summary table in the study includes the average value, standard deviation, minimum value, and maximum value of each variable. The thematic map of each variable was classified using the Natural Breaks method with three classes. The natural breaks classification method is a grouping of data based on the natural characteristics of the data [19]. This method can reduce variation between data within a group and increase variance between groups.

ICDSOS The 3rd International Conference on Data Science and Official Statistics November 27 - 28, 2025

Table 2. Summary Statistics.

Variable	Mean	SD	Min	Max
TBC cases	1570.52	2989.77	49	17161
Population density	1159.93	2221.83	41.16	8902.16
Per Capita Expenditure	11317.88	2104.39	6740	16069
Percentage of Health Fund Allocation	20.53	7.53	11	47.06
Mean Years of Schooling	9.50	1.32	6.4	11.82
Percentage of Smoking Population	21.75	4.43	12.08	26.86
Rainfall	2879.79	468.01	2124.81	3737.59
Number of Health Facilities	53.09	57.25	12	308
Number of Toddlers Vaccinated with BCG	4626	4943.73	450	23485

Based on the statistical summary of North Sumatra Province above, the average number of TB cases at the district/city level reached 1570 to 1571 cases, with a high variation of up to 2989.77 cases, while the region with the smallest number of TB cases was only 49 cases. This value is quite far below the average value, indicating the existence of outlier regions, which are likely small districts/cities. The population density in North Sumatra Province averages 1160 people/km², but there are areas with extreme values of up to 8902 people/km², indicating the existence of outliers in large urban areas. Next, the per capita expenditure of residents in North Sumatra averages Rp11,317,880/capita/year, with a reasonably small data distribution of Rp2,104,390/capita/year, indicating that the per capita expenditure of North Sumatra residents is not very different. The allocation of The Regional Revenue and Expenditure Budget funds to the health sector averages 20.53%, with the highest allocation being 47.06%. This figure is significantly higher than the average value and thus indicates an outlier. The average length of schooling is 9 to 10 years, with the highest average length of schooling being 11 to 12 years, indicating that there are still many residents of North Sumatra who have not fulfilled the Indonesian compulsory education period, which is 12 years. The average percentage of the population who smoke is 21.75%, with a distribution of 4.43%, indicating a low distribution. The average number of residents who smoke is less than ¼ of the population, indicating that not many residents smoke. The average rainfall in North Sumatra reaches 2879.79 mm/year. This value is included in the moderate rainfall category according to the Indonesian Agency for Meteorology, Climatology, and Geophysics, rainfall classification in Ruqoyah et al. (2023). The number of health facilities in the Regency/City of North Sumatra ranges from 12 to 308 units. This indicates inequality because the difference between the region with the lowest number of facilities and the highest is quite significant. Furthermore, the average number of toddlers vaccinated with BCG was 4,626, with a data distribution of 4,943,733. This figure is quite prominent, indicating that the number of infants vaccinated with BCG in North Sumatra still experiences significant spatial disparities. Based on these findings, further verification using thematic maps was conducted to examine the spatial condition of the data.

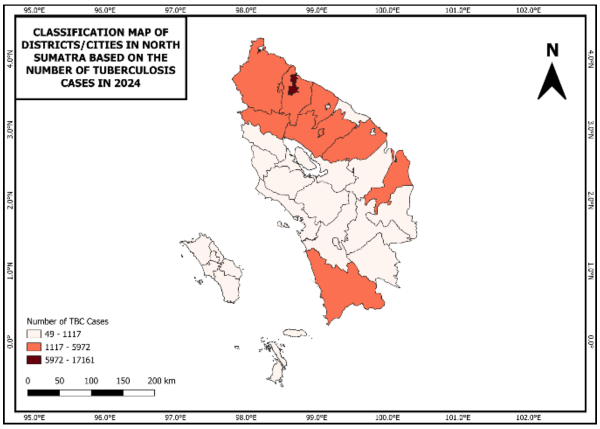

Figure 3. TBC Cases (Y) (Cases).

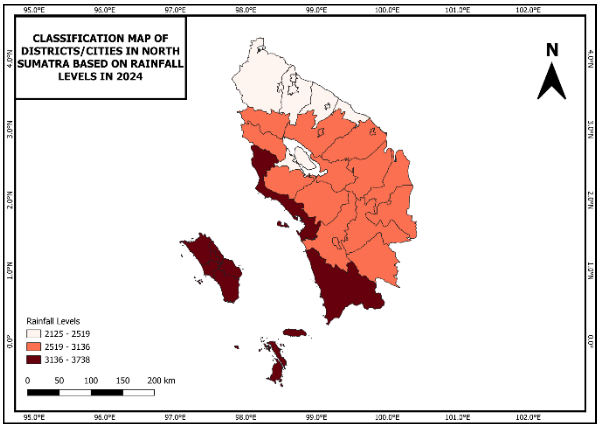

Figure 4. Rainfall (X_1) (mm).

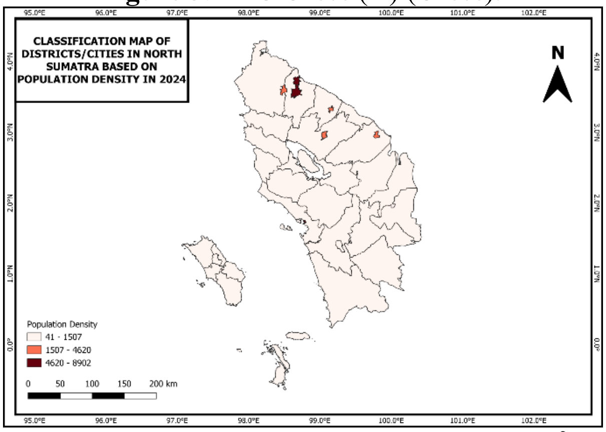

Figure 5. Population Density (X_2) (people/km^2).

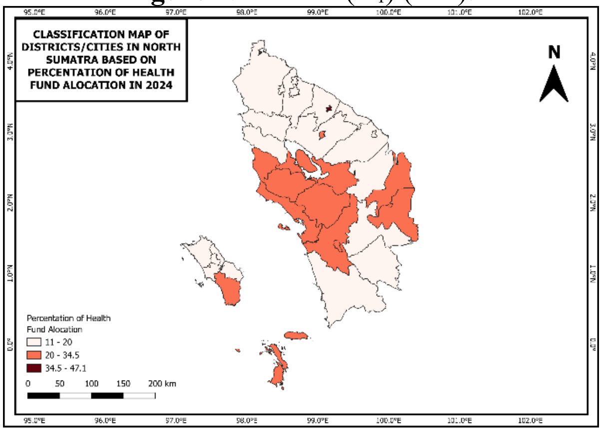

Figure 6. Health Fund Alocation (X_3) (%).

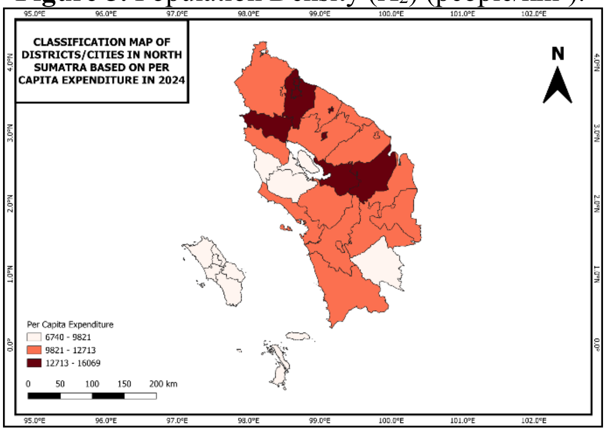

Figure 7. Per Capita Expenditure (X_4) (Thousand Rupiah/Capita).

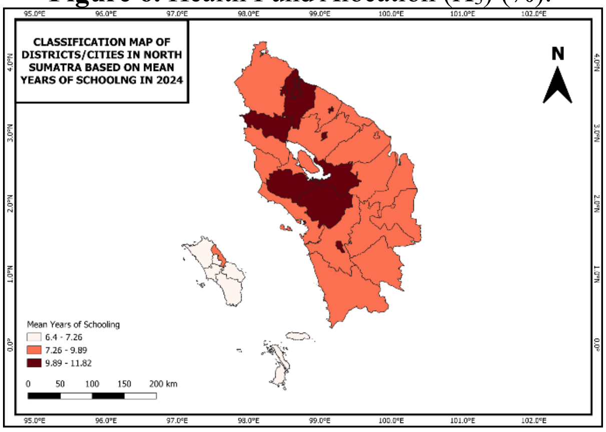

Figure 8. Mean Years of Schooling (X_5) (Years).

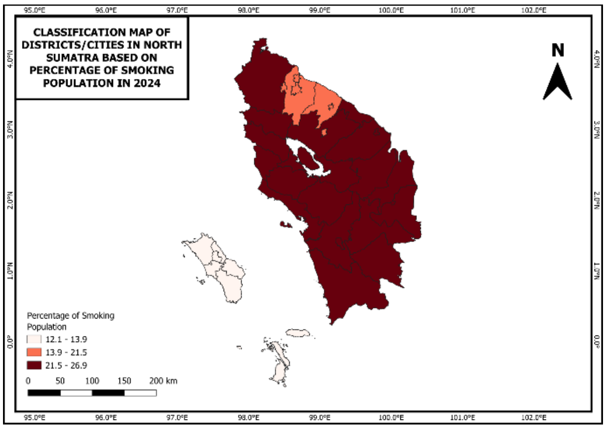

Figure 9. Percentage of Smoking Population (X_6) (%).

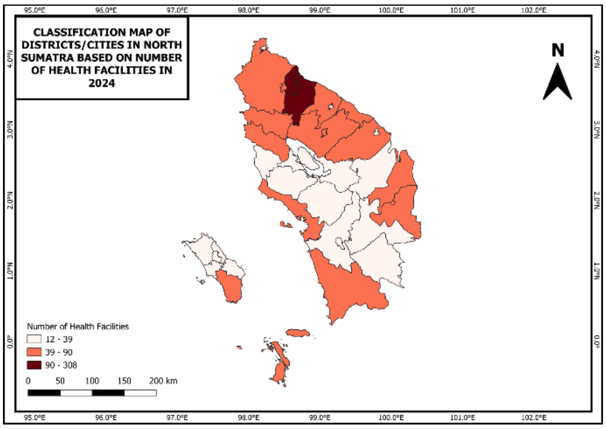

Figure 10. Number of Health Facilities (X_7) (Unit).

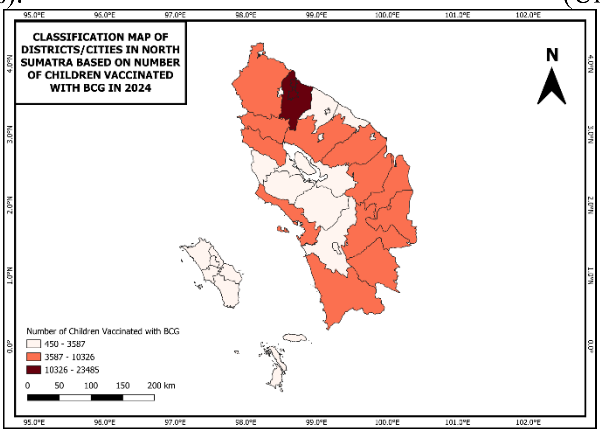

Figure 11. Number of Children Vaccinated with BCG (X_8) (People).

Figure 3 illustrates that, within the province of North Sumatra, the regions of Deli Serdang Regency and Medan City reported as the highest counts of tuberculosis infections, with 2,967 and 2,697 instances in the year 2024, respectively. These two administrative areas share a border and are neighbored by other districts experiencing intermediate levels of tuberculosis prevalence. Conversely, Nias Regency has the smallest number of tuberculosis infections with 33 cases in 2024. Nias Regency is geographically separate from the main island of Sumatra and is adjacent to several other districts that exhibit minimal occurrences of tuberculosis. The arrangement of tuberculosis infections across North Sumatra Province appears to form clusters, suggesting a spatial correlation among the cases. Specifically, two regencies or cities are identified within the zones showing the greatest tuberculosis case numbers, six regencies or cities exist within areas characterized by moderate tuberculosis case numbers, and twenty-five regencies or cities are situated in areas marked by low tuberculosis case numbers.

Figure 4 presents data about rainfall conditions in North Sumatra Province during 2024. Minahasa and its surrounding areas experienced high volumes of rainfall. Medan City and Deli Serdang Regency, with the highest tuberculosis cases in North Sumatra, generally experience low volumes of rainfall. According to research by Rolleh et al. (2025), rainfall is acknowledged as one of the factors influencing TB incidence.

Figure 5 shows that most regions in North Sumatra Province are classified as having low population density. Furthermore, Medan, Sibolga, and Tebing Tinggi are the only areas with high population density. Meanwhile, Binjai, Tanjungbalai, Padang Sidempuan, and Pematangsiantar are among the areas recognized as having medium population density. Medan, which also records the highest incidence of TB cases, is characterized by a high population density, thereby suggesting a direct association between these variables.

Figure 6 details the percentage distribution of funding allocations for healthcare across North Sumatra Province. It shows that only one district, Tebing Tinggi Regency, showed a substantial proportion of its budget dedicated to healthcare, accounting for 47.06%. This shows a strategic emphasis on health sector development in this regency. However, the map further reveals that a significant number of regencies and cities in North Sumatra Province continue to allocate only minimal financial resources to healthcare.

Figure 7 depicts the per capita expenditure levels observed throughout North Sumatra. In most parts of the province, residents generally have moderate levels of per capita expenditure, while the other areas demonstrate elevated levels. This implies a typically solid economic landscape in North Sumatra Province. The regions characterized by high per capita spending encompass Toba Samosir, Tebing Tinggi, Pematangsiantar, North Labuanbatu, Medan, Karo, and Deli Serdang. In this context, Medan and Deli Serdang, two regions distinguished by high rates of tuberculosis, are also identified within the high per capita spending bracket.

Figure 8 illustrates how the Average Years of Schooling (ALS) are spread across the different Regencies/Cities in North Sumatra. The visual indicates that a majority of the areas in North Sumatra, precisely 19 regencies/cities, are in the medium ALS range, marked by an educational attainment of 7 to 9 or 10 years. Eleven areas exhibit a high ALS level, namely Toba Samosir, Tebingtinggi, North Tapanuli, Sibolga, Pematangsiantar, Padangsidempuan, Medan, Binjai, Karo, Humbang Hasundutan, and Deli Serdang. Regarding this specific educational metric, two areas known for significant TB cases also demonstrate a high ALS, registering between 9 to 11 or 12 years of formal education. Nevertheless, considering Indonesia's mandate for 12 years of compulsory education, this is not regarded as satisfactory.

Figure 9 presents a breakdown of the percentage of smokers in each regency/city within North Sumatra. The majority of areas in North Sumatra have a significant proportion of smokers, with percentages ranging from 21.5 to 26.9% of the total population in each of these regencies/cities. The proportion is considerably high considering that approximately a quarter of the population in these locations still engage in smoking, a habit that can weaken their respiratory systems. The 23 areas marked by elevated smoking rates are Mandailing Natal, Langkat, North Labuanbatu, South Labuanbatu, Labuanbatu, Tanjungbalai City, Karo, Humbang Hasundutan, Dairi, Asahan, Batu Bara, Padang Lawas, North Padang Lawas, Padang Sidempuan, Pakpak Bharat, Pematangsiantar, Samosir, Sibolga, Simalungun, South Tapanuli, Central Tapanuli, North Tapanuli, and Toba Samosir.

Figure 10 indicates that only three regions have a high number of health facilities, ranging from 90 to 308. These three regions are Padangsidempuan, North Labuanbatu, and Asahan. The count of areas with medium and low health facility numbers is evenly split, with 15 in each category. The two regions with high TB cases fall into different categories: Medan, with 39 health facilities, and Deli Serdang, with 19 health facilities in the low category.

Figure 11 presents the number of infants vaccinated with BCG, a crucial immunization that supports the body's defense mechanisms against the tuberculosis virus. This vaccine is critical for ensuring long-term health and wellness. However, only a select few areas show a high number of infants being vaccinated with BCG. The province of North Sumatra predominantly exhibits moderate vaccination rates. Only two regions fall into the high category: Medan City and Deli Serdang, both of which also struggle with high tuberculosis incidence. In light of this, factors besides the BCG vaccine are probably contributing to the elevated tuberculosis rates in Medan City and Deli Serdang.

3.2. Inferential Analysis

3.2.1. Nonmulticolinearity Check

Before performing Poisson regression modeling, it is necessary to first check for multicollinearity to ensure that the independent variables are not highly correlated with one another. The Variance Inflation Factor (VIF) values for each independent variable are shown in Table 3.

Table 3. VIF value of predictor variables

Variable	VIF
Rain Intensity	2.456200
Population Density	2.558523
Government Health Spending	1.258478
Per Capita Expenditure	5.647933
Average Years of Schooling	3.631376
Percentage of Smokers	2.502726
Health Facilities	2.377451
Percentage of BCG	1.403376

The results of the multicollinearity check show that the VIF value of each independent variable does not exceed the maximum VIF limit, which is 10 [20]. This indicates that there is no significant correlation between the independent variables. Therefore, the non-multicollinearity assumption is satisfied, permitting the inclusion of all independent variables in the analysis.

3.2.2. Testing the Equidispersion Assumption

To analyze the relationship between the independent and dependent variables, Poisson regression was applied, as the dependent variable consists of count data. This type of regression requires the equidispersion assumption to be satisfied. Based on the results of the equidispersion assumption test, a p-value of 1.547e-05 and a dispersion value of 93.82736 were obtained, indicating the presence of overdispersion at the 5 percent significance level. Therefore, the Poisson regression model is not appropriate for use. Forcing the application of Poisson regression may lead to misinterpretation of the estimated standard deviation and standard errors, which can affect the significance of the parameter estimates. Overdispersion in count data can be addressed by using the negative binomial regression model. This model includes a dispersion parameter, making it suitable for handling overdispersion.

3.2.3. Spatial Autocorrelation Testing

Spatial autocorrelation testing was conducted to determine whether the number of tuberculosis cases in a given area is associated with the number of cases in neighboring regions. In this study, Moran's I test was used with Queen Contiguity weights. Based on the randomization results with 999 permutations, a Moran's I index value of 0.3495 and a p-value of 0.001 were obtained. These results indicate the presence of positive spatial autocorrelation in the number of tuberculosis cases in North Sumatra Province.

3.2.4. Spatial Heterogeneity Testing

A spatial heterogeneity test was conducted to examine whether there is spatial variation in the number of tuberculosis cases in North Sumatra and in the variables suspected to influence them. The test was performed using the *Breusch-Pagan* method. The test results showed a test statistic value of 23.006, which is greater than the critical value of $\chi^2_{(0.05;8)} = 15.507$, with a p-value of 0.003356, which is less than 0.05. The results suggest that spatial heterogeneity exists across regions, making the negative binomial regression model inappropriate for this context. To properly address both spatial variation and overdispersion in the data, the Geographically Weighted Negative Binomial Regression (GWNBR) model is a more suitable choice.

3.2.5. Modeling with Geographically Weighted Negative Binomial Regression (GWNBR)

To apply the Geographically Weighted Negative Binomial Regression (GWNBR) model, a spatial weighting matrix is required. This matrix is constructed by determining the optimal bandwidth, which is obtained by minimizing the cross-validation (CV) score.

Table 5. Selection of optimal bandwidth and the best kernel function

Kernel Function	Optimum Bandwidth	CV Score
Adaptive Bi-Square	7.790244	7447233325
Adaptive Gaussian	2.999561	58130130
Fixed Bi-Square	1.677916	75868371
Fixed Gaussian	11.32643	100920425

Table 5 shows that the model with the Adaptive Gaussian kernel function produced the lowest CV score, suggesting that it performs better than the other models compared. The model uses the kernel *Adaptive Gaussian* with a bandwidth of 2.999561, meaning each location considers approximately three nearest neighbors in estimating local parameters. Although the number of neighbors remains constant, the coverage radius varies depending on spatial density. After the spatial weighting matrix is formed, modeling is carried out using GWNBR. A total of 33 GWNBR models will be generated, one for each of the 33 districts/cities in North Sumatra Province, resulting in 297 estimated parameter coefficients.

After the model was formed, simultaneous and partial tests were conducted. The simultaneous test was conducted by comparing the deviance value of the GWNBR model, which was 4357.515, to the $\chi^2_{(0.05;8)} = 15.507$. Thus, at a significance level of 5 percent, there is sufficient evidence to state that at least one parameter has a significant effect on the number of tuberculosis sufferers in North Sumatra Province in 2024. Next, a partial test was carried out by comparing the calculated z-value to the $Z_{\alpha/2} = 1.96$. If the z-value is greater than 1.96, then the variable has a significant partial effect on the number of tuberculosis. Based on the partial test results, 8 regional groups were obtained based on the similarity of variable significance, as shown in Table 6.

Table 6. Grouping of districts/cities based on significant variables

No.	Regency/City	Significant Variables
1	Nias, Nias Barat, Kota Gunungsitoli	X1 X3 X4 X5 X6 X7 X8
2	Mandailing Natal, Tapanuli Selatan, Tapanuli Utara, Toba Samosir, Labuhan Batu, Simalungun, Dairi, Karo, Deli Serdang, Pakpak Bharat, Samosir, Batu Bara, Padang Lawas Utara, Padang Lawas, Labuhan Batu Selatan, Labuhan Batu Utara, Kota Sibolga, Kota Tanjung Balai, Kota Pematang Siantar, Kota Medan	X1 X2 X4 X5 X6 X7 X8
3	Tapanuli Tengah	X1 X2 X4 X5 X6 X8
4	Asahan, Kota Padangsidimpuan	X1 X2 X4 X5 X6 X7

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

5	Langkat, Serdang Bedagai, Kota Tebing Tinggi, Kota Binjai	X1 X2 X3 X4 X5 X6 X7 X8
6	Nias Selatan	X1 X2 X4 X8
7	Humbang Hasundutan	X1 X2 X4 X6 X8
8	Nias Utara	X1 X4 X5 X6 X7 X8

Figure 12 shows a thematic map that groups districts/cities with the same combination of significant variables into a single color. This suggests that the factors contributing to tuberculosis cases vary across regions, highlighting the need for region-specific policies that address the relevant influencing variables.

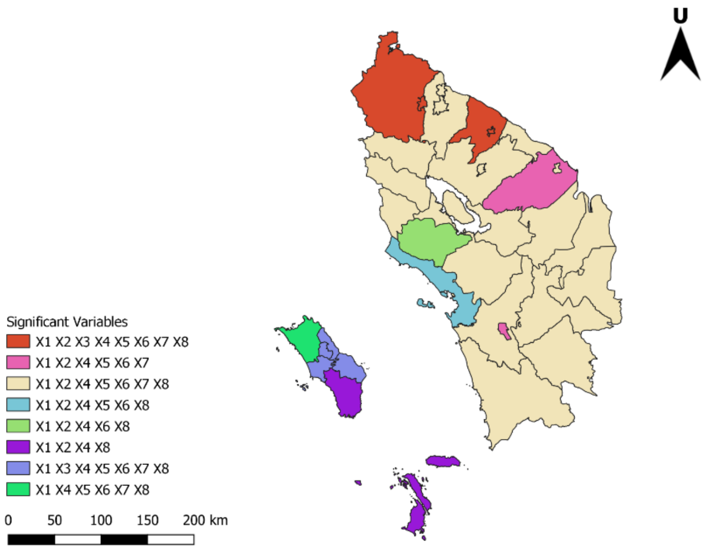

Figure 12. Significance map of variables influencing tuberculosis cases in North Sumatra..

For example, the parameter estimation shown in Table 7 was carried out to identify the significant variables that influence tuberculosis cases in North Nias Regency. The results of this estimation were then interpreted to understand the relationship between these variables and the number of tuberculosis cases.

Table 7. GWNBR model parameter estimation for North Nias Regency

	Coefficient	z-value
(Constant)	4.6014×10^{-7}	-1.8084×10^{11}*
X1	3.4011×10^{-4}	14.0370*
X2	2.9422×10^{-5}	-0.5823
X3	1.7286×10^{-4}	1.6396
X4	-1.4783×10^{-5}	-1.9041×10^{2}*
X5	-6.1131×10^{-7}	-2.1986×10^{7}*
X6	-9.1032×10^{-6}	20.4728*
X7	6.0023×10^{-4}	-2.9479*

X8	-5.3188 $\times 10^{-5}$	-1.3871 $\times 10^{3}$*

Note: *) Significant at the 5% significance level

Rain intensity has a statistically significant effect on tuberculosis incidence in North Nias Regency, with an estimated coefficient of 3.4011 $\times 10^{-4}$. This means that for every one-unit increase in rainfall, the number of tuberculosis cases in North Nias is expected to increase by $exp(3.4011 \times 10^{-4}) = 1.00034$ times, assuming other variables remain constant. Higher levels of rainfall intensity may alter ambient temperature and humidity, creating conditions that favor the survival of *Mycobacterium tuberculosis* and increase the risk of transmission [8].

Per capita expenditure shows a significant negative association with tuberculosis incidence in North Nias Regency, with an estimated coefficient of -1.4783 $\times 10^{-5}$. This implies that each one-unit increase in per capita expenditure is expected to reduce tuberculosis cases by 0.999985 times, assuming all other variables are held constant. This result aligns with the findings of Dlatu et al. [12], who reported that higher per capita expenditure is significantly associated with lower tuberculosis incidence. Individuals with greater financial resources are more likely to maintain proper household sanitation, secure safe drinking water, consume nutritious food, and afford healthcare services [21].

The average years of schooling variable has a significant effect on tuberculosis cases in North Nias Regency, with a coefficient of -6.1131 $\times 10^{-7}$. This means that for every one-year increase in average years of schooling, the number of tuberculosis cases is expected to decrease by $exp(-6.1131 \times 10^{-7}) = 0.99999$ times, assuming other variables remain constant. This finding is consistent with the study by [13], which found a negative relationship between average years of schooling and tuberculosis incidence. Both ecological and individual-level analyses show that the higher the average education level in a country, the lower the tuberculosis incidence rate.

The percentage of smokers has a significant effect on tuberculosis cases in North Nias Regency, with a coefficient of -9.1032 $\times 10^{-6}$. This means that for every 1% increase in the smokers population, the number of tuberculosis cases is expected to decrease by a factor of $exp(-9.1032 \times 10^{-6}) = 0.999991$, assuming other variables remain constant. This finding is inconsistent with the results of [10], which stated that smoking habits can increase the risk of tuberculosis.

The number of healthcare facilities significantly affects tuberculosis cases in North Nias Regency, with a coefficient of 6.0023 $\times 10^{-4}$. This means that for every one-unit increase in the number of healthcare facilities, tuberculosis cases are expected to increase by a factor of 1.0006, assuming other variables remain constant. This suggests that an increase in healthcare facilities does not automatically lead to a reduction in tuberculosis cases.

The percentage of BCG immunization significantly affects the number of tuberculosis cases in North Nias Regency, with a coefficient of -5.3188 $\times 10^{-5}$. This means that for every one percentage point increase in BCG immunization coverage, the number of tuberculosis cases is expected to decrease by a factor of $exp(-5.3188 \times 10^{-5}) = 0.999947$, assuming other variables remain constant. This result is consistent with the findings of [15], which reported that BCG immunization has an effectiveness rate of 85% and has been proven effective in preventing tuberculosis in children. Children who do not receive the BCG vaccine are 6.87 times more likely to develop childhood TB compared to those who are immunized.

4. Conclusion

Based on the results and discussion, the conclusions of this study are as follows.

1. The distribution of tuberculosis cases in North Sumatra Province in 2024 tended to be clustered and varied considerably between districts/cities. The data also showed spatial influence, as

evidenced by the results of the dependency and spatial heterogeneity tests, which showed significant results. This means that there is regional dependence and diversity between districts/cities in North Sumatra Province related to the number of tuberculosis cases.

2. GWNBR estimates show that the independent variables significantly influencing tuberculosis cases vary across districts/cities in North Sumatra, forming eight clusters. The variables that consistently have a significant influence in each district/city in North Sumatra are rain intensity and per capita expenditure. This suggests that rainfall and per capita expenditure are key determinants of tuberculosis cases across the province. It also highlights that both environmental and economic conditions contribute significantly to the number of tuberculosis in North Sumatra.

Based on the findings, several recommendations can be proposed as follows.

1. For the government, tuberculosis management policies should consider variables that significantly influence tuberculosis in the relevant region. Given that rainfall and per capita expenditure have been found to consistently affect TB incidence across all districts, these factors should be prioritized in designing targeted interventions and preventive measures.
2. For further research, methods that consider the influence of time, such as panel data analysis, can be used, as well as individual data analysis to incorporate additional variables such as age and gender. Furthermore, the use of primary data can also be considered to obtain more in-depth and accurate information.

References

[1] Komisi Nasional Hak Asasi Manusia, “Kerangka Analisis untuk Mengintegrasikan Pembangunan Berkelanjutan (SDGs) dengan Kewajiban Pemenuhan Hak-hak Asasi Manusia untuk Indonesia”

[2] World Health Organization, “Fact sheets – Tuberkulosis,” WHO Indonesia, accessed Jul. 16, 2025. [Online]. Available: https://www.who.int/indonesia/news/campaign/tb-day-2022/fact-sheets

[3] TB Indonesia, “UMUM_PNPK_revisi,” TB Indonesia, Jun. 2021. [Online]. Available: https://www.tbindonesia.or.id/wp-content/uploads/2021/06/UMUM_PNPK_revisi.pdf. [Accessed: Jul. 16, 2025].

[4] World Health Organization, “Global Tuberculosis Report 2024”. Geneva: WHO, 2025.

[5] Kementerian Kesehatan RI, “Indonesia's Movement to End TB,” 11 Apr. 2025.

[6] TB Indonesia, “Laporan Kegiatan Pertemuan Nasional Program Tuberkulosis Tahun 2024,” Pustaka TBC – TB Indonesia, May 9, 2025.

[7] H. L. Blum, "Evaluating Health Care," *Medical Care*, vol. 12, no. 12, pp. 999–1011, Dec. 1974. [Online]. Available: https://www.jstore.org/stable/3763529

[8] D. Gould and C. Brooker, *Mikrobiologi Terapan untuk Perawat*. Jakarta: Buku Kedokteran EGC, 2003.

[9] M. D. Kasaluhe, *Tuberkulosis: Pemetaan sebaran kasus melalui pemanfaatan sistem informasi geografis*. Jakarta: Penerbit NEM, 2021.

[10] M. R. Rosyid and A. M. Sakufa, "Hubungan kondisi fisik rumah dan kebiasaan merokok dengan kejadian tuberculosis di wilayah kerja Puskesmas Banjarejo Kota Madiun," *Jurnal Ilmu Kesehatan*, vol. 11, no. 2, pp. 76–94, June 2023. [Online]. Available: http://www.ejurnaladhkdr.com/index.php/jik/article/view/457. doi: 10.32831/jik.v11i2.457

[11] F. A. Költringer, K. S. Annerstedt, D. Boccia *et al*., "The social determinants of national tuberculosis incidence rates in 116 countries: a longitudinal ecological study between 2005–2015," *BMC Public Health*, vol. 23, no. 337, 2023. [Online]. Available: https://doi.org/10.1186/s12889-023-15213-w

[12] N. Dlatu, B. Longo-Mbenza, and T. Apalata, "Predictors of tuberculosis incidence and the effects of multiple deprivation indices on tuberculosis management in OR Tambo district over a 5-year period," *PLOS ONE*, vol. 17, no. 3, p. e0264811, Mar. 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0264811

[13] K. Machiyama, J. C. Semenza, R. J. Silverwood *et al*., "School life expectancy and risk for tuberculosis in Europe," *Int. J. Public Health*, vol. 61, pp. 603–611, 2016. [Online]. Available: https://doi.org/10.1007/s00038-015-0768-6

[14] C. Nisa, M. N. Aidi, and I. M. Sumertajaya, "Geographically weighted negative binomial regression modeling of tuberculosis cases with distribution evaluation," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 7, no. 4, pp. 279–285, Jul.–Aug. 2020. [Online]. Available: https://doi.org/10.32628/IJSRSET1207473

[15] A. F. Wulanda and S. Delilah, "Efektivitas imunisasi BCG terhadap kejadian tuberkulosis anak di Kabupaten Bangka (Effectiveness of BCG Immunization Against Children’s Tuberculosis Incidence in Bangka Regency)," *Jurnal Kesehatan Poltekkes Kemenkes Pangkalpinang*, vol. 9, no. 1, pp. 37–41, June 2021.

[16] T. Mumtaz and A. P. Utomo, "Modelling the number of new pulmonary tuberculosis cases with geographically weighted negative binomial regression method," *Indonesian Journal of Statistics and Its Applications*, vol. 2, no. 2, pp. 77–92, 2018.

[17] - Rizki, "Literature review: faktor risiko lingkungan kejadian tuberkulosis (Environmental Risk Factors of Tuberculosis Incidence)," Journal of Public Health Innovation, vol. 4, no. 2, pp. 476–483, June 2024.

[18] D. A. Mustika and R. Nooraeni, "Kajian Efek Spasial Kasus Difteri dengan Geographically Weighted Negative Binomial Regression (GWNBR)", *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 1, pp. 91-104, 2019.

[19] Esri, "Data classification methods," *ArcGIS Pro Documentation*, Jul. 16, 2025. [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm

[20] M. H. Kurtner, C. J. Nactsheim and J. Neter, "Multiple Linear Regression," *Applied Linear Regression Model*, Fourth Edition, New York: McGraw-Hill/Irwin, 2004, 409.

[21] I. Yuniar, S. Sarwono, and S. D. Lestari, "Hubungan Status Gizi dan Pendapatan terhadap Kejadian Tuberkulosis Paru," *Jurnal Perawat Indonesia*, vol. 1, no. 1, pp. 18-25, 2017.I.

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025