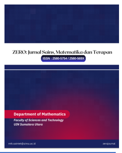
Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 1, 2025
DOI: 10.30829/zero.v9i1.24877
Page: 239-246



Lasso Quantile Regression for Predictive Modeling of Dengue Hemorrhagic Fever Incidence in Indonesia

Samsul Arifin
Department of Statistics, Lambung Mangkurat University, Banjarbaru, Indonesia

Dewi Anggraini
Department of Statistics, Lambung Mangkurat University, Banjarbaru, Indonesia

Dewi Sri Susanti
Department of Statistics, Lambung Mangkurat University, Banjarbaru, Indonesia

Article Info

Article history:

Accepted, 07 05 2025

Keywords:

Dengue Hemorrhagic Fever (DHF);
LASSO;
Multicollinearity;
Quantile Regression;

ABSTRACT

Dengue Hemorrhagic Fever (DHF) is an endemic disease in Indonesia, marked by uneven distribution patters driven by environmental, social, and economic factors. This study develops a predictive model for DHF incidence using the LASSO Quantile Regression approach, which estimates conditional quantiles (τ = 0.25, 0.50, 0.75) by minimizing the check loss function with an added L1 penalty term. This method enables variable selection and captures non-uniform effects across quantiles while addressing multicollinearity. The data used includes nine predictor variables obtained from BPS for the year 2025 and were analyzed using R version 2024.12.0. The results show that urban/rural area size significantly influences all quantiles, while poverty rate and number of healthcare facilities are significant only at τ = 0.25 and τ = 0.50. The model achieves its best predictive performance at τ = 0.25 with a pseudo-R^2 of 0.2838. These findings demonstrate the model's robustness in capturing varying risk factors.

Corresponding Author:

Samsul Arifin,
Department of Statistics,
Lambung Mangkurat University, Banjarmasin, Indonesia
Email: samsularrr@ulm.ac.id

1. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) remains one of the major public health problems in Indonesia [1]. The disease is transmitted by Aedes aegypti mosquitoes and often occurs seasonally, with incidence rates varying across regions [2]. Data from the Ministry of Health indicate that the number of DHF cases in Indonesia fluctuates significantly each year, highlighting the need for accurate modeling to understand the factors influencing its spread [3].

Previous studies have applied various statistical methods to analyze the determinants of DHF incidence, including multiple linear regression [4], logistic regression [5], and Poisson regression [6]. However, these approaches mainly capture the average effects of predictor variables, making them less effective in identifying different relationship patterns at lower and upper quantiles. Additionally, Poisson regression assumes overdispersion where the mean and variance of count data are equal [7] . This can lead to underestimated standard errors and biased statistical inference. The incidence of DHF is unevenly distributed and influenced by different risk factors at lower, median, and upper ends of the distribution [8], which calls for more flexible analytical frameworks.

Quantile regression provides a robust alternative by estimating conditional quantiles of the response variable, denoted by τ, where $\tau \in (0, 1)$. For example, τ = 0.25, 0.50, and 0.75 represent the 25 percentile, 50 percentile as median, and 75 percentile of DHF incidence, respectively [9]. These quantiles were chosen to capture low, moderate, and moderately high levels of incidence, offering a balanced view of risk without the estimation instability often encountered at extreme quantiles such as τ = 0.05 or 0.95 [10]. Quantile regression is also more robust in the presence of outliers and is suitable when data do not follow a normal distribution [11].

In addition, DHF incidence is influenced by many environmental, demographic, and socio-economic factors that are often correlated with each other [12]. To manage this complexity, regularization techniques such as the LASSO can be applied. LASSO helps to select the most relevant predictors, reduce multicollinearity [13], and prevent overfitting, which improves model accuracy and interpretation [14]. Combining quantile regression with LASSO creates a powerful and efficient modeling framework that can handle complex and diverse data while still giving reliable predictions [15]. This approach is especially useful in public health, where decision-makers need to understand risks across different levels of severity.

Importantly, this method also has real-world benefits. Because the pattern and risk factors of DHF vary across regions, quantile-based models provide detailed insights that can help local governments identify high-risk areas and take more targeted action. The results can support better allocation of resources, more effective prevention strategies, and stronger local disease surveillance. Based on this background, the aim of this study is to develop a quantile regression model with LASSO regularization to examine the factors that influence DHF incidence in Indonesia in 2024. This method addresses previous limitations and supports more adaptive, evidence-based public health strategies at the regional level.

2. RESEARCH METHOD

Quantile Regression

Quantile Regression was introduced by Koenker and Bassett (1978) as a generalization of the classical linear regression model [16]. The quantile regression estimates the conditional quantiles of the response variable, denoted as $Q_Y(\tau \mid X)$, where $\tau \in (0,1)$ represents the quantile level of interest commonly 0.25 as a lower quartile, 0.50 as a median, or 0.75 as a upper quartile, while Ordinary Least Squares (OLS) regression estimates the conditional mean of the dependent variable, denoted as $E(Y|X)$ [17]. This approach provides a more comprehensive analysis of the conditional distribution of Y, capturing heterogeneity across the distribution that OLS might overlook.

In this study, the quantile levels τ = 0.25, 0.50, and 0.75 were selected to reflect different degrees of DHF incidence. Specifically, τ = 0.25 captures the behavior of predictors in areas with relatively low incidence, τ = 0.50 represents the median condition, and τ = 0.75 focuses on moderately high incidence. This selection allows the model to explore whether the influence of covariates varies across different levels of disease burden. The use of these central quantiles ensures robustness against outliers while still providing meaningful interpretation across the distribution without concentrating only on the extremes, which may be more sensitive to noise in the data.

The estimation steps of quantile regression are outlined as follows:

1. Model Specification
   Let $Y_i \in R$ denote the dependent variable, $X_i \in R^p$ the vector of covariates, and $\beta_\tau \in R^p$ the regression parameters at quantile level τ. The quantile regression model is specified as:
   $$Q_Y(\tau \mid X_i) = \boldsymbol{X_i^T \beta_\tau} \quad (1)$$
2. Defining the Loss Function
   Quantile regression employs the check loss function, which is defined as:
   $$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases} \quad (2)$$
   where $u = y_i - \boldsymbol{X_i^T \beta}$. This function applies asymmetric weights to residuals above and below the target quantile, allowing the model to capture conditional quantile behavior effectively.
3. Parameter Estimation via Optimization
   To estimate the quantile regression coefficients $\hat{\beta}_\tau$, numerical optimization techniques are applied [18]. These typically involve linear programming methods such as the Simplex algorithm or Interior Point Method, depending on the size and complexity of the data.
   $$\hat{\beta}_\tau = \arg \min_{\beta \in R^p} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{X_i^T \beta}\right) \quad (3)$$
4. Model Evaluation
   Model performance in quantile regression can be assessed using the R-Squared Pseudo [19], calculated as follows:
   $$R^2_{pseudo} = 1 - \frac{\sum_{i=1}^{n} \rho_\tau(y_i - X_i^T \hat{\beta}_\tau)}{\sum_{i=1}^{n} \rho_\tau(y_i - \hat{y}_\tau)} \quad (4)$$

LASSO Regression

LASSO regression is a regularization method used to simplify predictive models through automatic variable selection and to prevent overfitting, especially when the number of predictors is large or when multicollinearity is present. LASSO achieves this by shrinking some regression coefficients exactly to zero, thereby producing a more parsimonious or sparse model [20]. The objective function of LASSO regression is defined as:

$$\hat{\beta} = \arg \min_{\beta \in R^p} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \boldsymbol{X}_i^T \boldsymbol{\beta} \right)^2 + \lambda ||\beta||_1 \right\} \quad (5)$$

Where $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$ represents the LASSO penalty function, and $\lambda \geq 0$ is the regularization parameter that controls the strength of the penalty [21].

LASSO Quantile Regression

LASSO Quantile Regression LASSO or penalized quantile regression, is an extension of the standard Quantile Regression that incorporates the LASSO penalty into its objective function [22]. This model extension is capable of addressing multicollinearity issues, performing automatic variable selection, and is well-suited for high-dimensional data or models with a large number of predictors. The estimation steps of LASSO quantile regression as follows:

1. Model Specification
   Let $Y_i \in R$ denote the dependent variable, $X_i \in R^p$ the vector of covariates, and $\beta_\tau \in R^p$ the regression parameters at quantile level τ. The conditional quantile function is a model as:
   $$Q_Y(\tau \mid X_i) = \boldsymbol{X}_i^T \boldsymbol{\beta}_\tau \quad (6)$$
2. Quantile Loss Function
   The quantile loss function $\rho_\tau(u)$, also known as the check function, is defined as:
   $$\rho_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ (\tau - 1)u & \text{if } u < 0 \end{cases} \quad (7)$$
   where $u = y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}$. This loss function applies asymmetric penalties, allowing the model to capture the effects of predictors at different quantiles of the outcome distribution.
3. Parameter Estimation with LASSO Regularization
   To enhance sparsity and reduce overfitting, a regularization term is added to the objective function. The LASSO-penalized quantile regression minimizes the following:
   $$\hat{\beta}_\tau = \arg \min_{\beta \in R^p} \left\{ \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{X}_i^T \boldsymbol{\beta} \right) + \lambda ||\beta||_1 \right\} \quad (8)$$
   $\lambda \geq 0$where is the regularization parameter controlling the degree of penalization? Higher λ values increase sparsity by shrinking less important coefficients toward zero. Optimization is commonly solved via coordinate descent or interior point methods, especially when embedded in high-dimensional settings. In this framework, any coefficient estimated as $\hat{\beta}_j = 0$ is considered to have been excluded from the model through LASSO variable selection process, indicating that the associated predictor is not statistically relevant at the given quantile level. This shrinkage mechanism simplifies the model while retaining only the significant variables [23].
4. Model Evaluation
   Model performance is evaluated using the R-Squared Pseudo [19], which measures the proportion of quantile loss explained by the model:
   $$R^2_{pseudo} = 1 - \frac{\sum_{i=1}^{n} \rho_\tau(y_i - X_i^T \hat{\beta}_\tau)}{\sum_{i=1}^{n} \rho_\tau(y_i - \hat{y}_\tau)} \quad (9)$$
   where $\hat{y}_\tau$ the predicted value from a baseline model (e.g., median or constant model). A higher R^2_{pseudo} indicates better model fit.

3. RESULT AND ANALYSIS

This study is quantitative research aimed at developing a predictive model for the number of DHF cases in Indonesia using the Quantile Regression method with a LASSO penalty function. Nine important predictor variables that are thought to affect the prevalence of DHF in Indonesia are included in the model. These variables are population density (x1), percentage of households with access to clean water (x2), percentage of households with proper sanitation facilities (x3), percentage of households using earthen floors (x4), percentage of the population living below the poverty line (x5), total area of urban or rural regions (x6), number of healthcare facilities such as community health centers and hospitals (x7), number of healthcare workers (x8), and climate-related factors including monthly average rainfall and temperature (x9). These variables were selected based on their relevance in epidemiological and environmental literature. The data were obtained from the publication of The Central Statistics Agency (BPS) Indonesia, *Indonesia in Figures 2025* [24]. Data were analyzed using R version 2024.12.0.

Data Visualization

The boxplot in Figure 1 is the visualization data to illustrate data distribution, central tendency, and to detect the presence of outliers for each variable [25]. Through the use of boxplots, the researcher can identify whether any variables exhibit asymmetric distributions, wide interquartile ranges, or extreme values that may affect parameter estimation in the model [26]. In the context of this study, boxplots were applied to the nine predictor variables, as shown in Figure 1. The presence of outliers, as revealed by the boxplots, serves as a key justification for the use of quantile regression, which is more robust to outliers compared to ordinary linear regression.

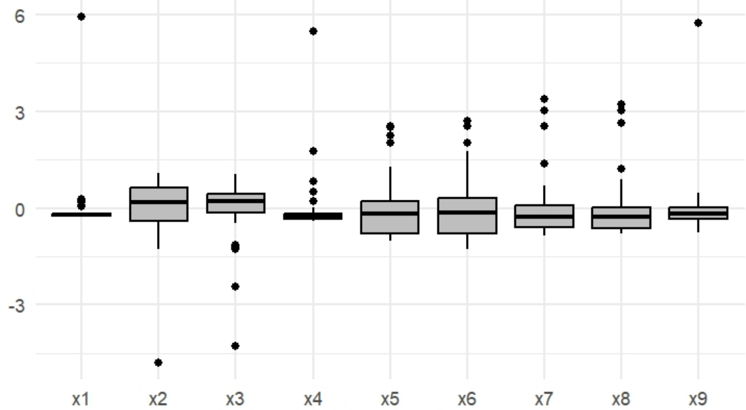


Figure 1. Boxplot visualization for outlier detection

Figure 1 presents the results of data visualization using boxplots. Variables x2, x3, and x4 display wider interquartile ranges compared to the other variables, and they also contain visible outliers. This indicates disparities in distribution across regions, such as unequal access to clean water and sanitation. Variable x4, which represents the use of earthen floors, is highly skewed to the lower end, with numerous extreme values, suggesting that only a small proportion of regions still have households with earthen flooring. Meanwhile, other variables such as x5 to x9, which cover socio-economic indicators, healthcare facilities, and environmental conditions, exhibit relatively more stable distributions but still show the presence of outliers. These findings highlight the significant presence of outliers in the dataset, thereby reinforcing the appropriateness of using quantile regression to model DHF cases in Indonesia.

Multicollinearity Testing

In regression modeling, particularly when involving multiple predictor variables, it is essential to ensure that there is no high degree of linear correlation among the variables, a condition known as multicollinearity [27]. High multicollinearity can lead to unstable parameter estimates, increased variance, and reduced accuracy in model interpretation [28]. Therefore, prior to model development, multicollinearity testing should be conducted using the Variance Inflation Factor (VIF) as an indicator. A high VIF value indicates a significant multicollinearity issue that must be addressed before proceeding with the modeling process.

Table 1. Multicollinearity testing

Variable	VIF Value	Result
Population Density (x1)	4.1896	Not Multicollinearity
Percentage of Households with Access to Clean Water (x2)	4.7320	Not Multicollinearity
Percentage of Households with Proper Sanitation (x3)	5.4236	Not Multicollinearity
Percentage of Households Using Earthen Floors (x4)	3.9382	Not Multicollinearity
Percentage of Population Living in Poverty (x5)	3.5413	Not Multicollinearity
Urban/Rural Area Size (x6)	1.4377	Not Multicollinearity
Number of Healthcare Facilities (x7)	43.4265	Multicollinearity
Number of Healthcare Workers (x8)	48.3952	Multicollinearity
Rainfall and Average Monthly Temperature (x9)	1.7161	Not Multicollinearity

The results of the multicollinearity test presented in Table 1 indicate that most variables have VIF values below the threshold of 10, suggesting no significant multicollinearity among these predictors. The highest VIF values were observed in the variables Number of Healthcare Facilities (x7) and Number of Healthcare Workers (x8), at 43.43 and 48.40, respectively, indicating very strong multicollinearity. This can be logically explained by the fact that regions with a higher number of healthcare facilities typically also have more healthcare personnel, leading to a strong correlation between the two.

The presence of multicollinearity in these two variables warrants careful consideration during the modeling process, as it may affect the stability of coefficient estimates. In this study, the use of the LASSO method is highly appropriate, as LASSO not only performs regularization to prevent overfitting but also enables automatic variable selection, addressing multicollinearity by retaining the more statistically dominant variable and eliminating the redundant one. Therefore, despite the existence of multicollinearity in some predictors, the adoption of a penalized regression approach provides a strategic solution to develop a more stable and accurate predictive model.

Estimation of the LASSO Quantile Regression Model

Quantile regression with a LASSO penalty function was employed in this study as a predictive approach capable of estimating not only the median quantile (τ = 0.50), but also the lower (τ = 0.25) and upper (τ = 0.75) quantiles of DHF case counts in Indonesia. The main advantage of this method lies in its ability to handle data heterogeneity and to perform automatic variable selection through penalization, which is particularly beneficial when dealing with potential multicollinearity or a large number of predictor variables. The model estimation results are presented in Table 1 and are further supported visually by coefficient plots for each quantile, as shown in Figure 2. This comparison allows researchers to assess the differential effects of each predictor variable across various quantile levels of dengue incidence.

Table 2. Estimation of the LASSO quantile regression model

Variable	Quantile		
	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
(Intercept)	0.1315	0	-0.2809
x1	0	0	0
x2	0	0	0
x3	0	0	0
x4	0	0	0
x5	-0.1050	-0.0915	0
x6	0.2532	0.2205	0.0726
x7	-0.0887	-0.0580	0
x8	0	0	0
x9	0	0	0

Based on the estimation results of the LASSO Quantile Regression model presented in Table 1, only three predictor variables percentage of poor population (x5), area size (x6), and number of health facilities (x7) were retained with non-zero coefficients at one or more quantiles. The remaining variables (x1, x2, x3, x4, x8, and x9) were automatically shrunk to zero by the LASSO penalty function across all quantiles. This indicates that these variables do not contribute significantly to the predictive power of the model in explaining DHF incidence and are therefore excluded to reduce complexity and prevent overfitting.

Setting the coefficients of these variables to zero reflects their lack of statistical contribution in the presence of stronger, more relevant predictors. In other words, the model treats these variables as irrelevant or redundant for the purpose of predicting DHF case counts across the distribution. This variable elimination is a central strength of the LASSO approach, ensuring that only meaningful predictors are retained, which improves model interpretability and robustness. The predictive models of DHF cases in Indonesia for each quantile are formulated as follows:

$$y(0.25) = 0.1315 - 0.1050\text{x5} + 0.2532\text{x6} - 0.0887\text{x7}$$
$$y(0.50) = -0.0915\text{x5} + 0.2205\text{x6} - 0.0580\text{x7}$$
$$y(0.75) = -0.2809 + 0.0726\text{x6}$$

At quantile τ = 0.25, the percentage of poor population (x5) and the number of health facilities (x7) exhibit negative effects, indicating that regions with more health facilities and higher poverty levels tend to have lower DHF incidence. This can be interpreted as an indication that in areas with low DHF case counts, poverty does not necessarily correlate with increased risk possibly due to other factors such as vector distribution or preventive behaviors. Conversely, the area size (x6) variable shows a positive coefficient, suggesting that larger regions tend to report more DHF cases even in the lower quantile, which may result from spatial population dispersion or challenges in vector control.

At the median quantile τ = 0.50, a similar pattern is observed: x5 and x7 still show negative coefficients, and x6 remains dominantly positive. However, in the upper quantile (τ = 0.75), only x6 retains a non-zero coefficient, while all other variables are eliminated by the LASSO penalty function. This indicates that when DHF case counts are very high, area size becomes the only consistently influential predictor. The negative intercept coefficient at τ = 0.75 also suggests that the baseline prediction tends to be lower when other predictors are inactive.

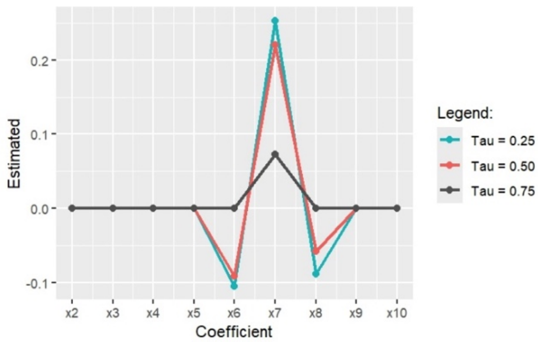


Figure 2. Plot of estimated coefficients across quantiles

The coefficient plot results in Figure 2 show that variable x6 consistently appears across all quantiles with a positive direction, although its magnitude tends to decrease as the quantile increases. This consistent significance of x6 may be attributed to population effects, as larger administrative areas are likely to encompass more people, thereby increasing the likelihood of reported dengue cases. On the other hand, variables x5 and x7 are only present in the lower and median quantiles but are eliminated in the upper quantile. This visualization highlights how the quantile regression approach enables the identification of variable contributions that are not homogeneous across the entire distribution, which is particularly important in the epidemiological context of DHF, where risk patterns may vary between regions with low and high incidence.

Overall, the results confirm that the LASSO quantile regression method not only enhances predictive accuracy but also assists in identifying which predictors matter most at specific points in the DHF incidence distribution. The elimination of irrelevant variables improves interpretability and provides a clearer picture of the key factors driving DHF in different regional and epidemiological contexts.

Evaluation of LASSO Quantile Regression Prediction Model

The predictive performance of the LASSO Quantile Regression model in this study is evaluated using R^2 pseudo values, which serve as measures of goodness-of-fit to indicate how well the model explains the variability of the data at each quantile level, as shown in Table 3.

Table 3. Model evaluation of LASSO quantile regression

Quantile	R^2 pseudo
$\tau = 0.25$	0.2838
$\tau = 0.50$	0.0477
$\tau = 0.75$	0.1887

The highest R^2 pseudo is observed at the lower quantile $\tau = 0.25$, with a value of 0.2838. This indicates that the model performs better in explaining the variation in dengue fever cases in regions with low incidence. This likely reflects more consistent and predictable relationships between predictor variables and DHF incidence at this level. In contrast, the R^2 pseudo values for the median $\tau = 0.50$ and upper quantile $\tau = 0.75$ are relatively low at 0.0477 and 0.1887, respectively suggesting that the model has limited explanatory power for regions with medium to high incidence rates. This may be due to the complex dynamics of DHF in these areas, where risk factors are likely influenced by unmeasured variables such as public behavior, population mobility, or the effectiveness of prevention programs.

Overall, these results confirm that the LASSO Quantile Regression model is more effective at the lower quantile than modeling the middle and upper quantiles. Therefore, further model development or the inclusion of more representative predictors may be necessary to enhance predictive performance for areas with higher case distributions.

4. CONCLUSION

This study demonstrates that the LASSO Quantile Regression model provides strong predictive performance for DHF cases in Indonesia, particularly in regions with low incidence ($\tau = 0.25$), as indicated by a high R^2 pseudo value of 0.2838. The area size variable consistently emerged as the primary predictor across all quantiles, while variables such as poverty rate and number of healthcare facilities only showed significance at the lower and middle quantiles. The LASSO approach proved effective in selecting relevant variables and addressing multicollinearity. However, the model's predictive accuracy declined at the median and upper quantiles, highlighting the need for further model refinement or additional variables to improve predictions in areas with moderate to high DHF incidence.This study is expected to provide strategic input for the government, both at the central and regional levels, in designing more data-driven and locally risk-based DHF prevention and control policies. In low-incidence areas, the model can serve as an early warning tool and a basis for more efficient resource allocation, such as deploying healthcare personnel, distributing larvicides, or conducting educational campaigns. Furthermore, since socio-economic variables like poverty and access to healthcare facilities have a significant influence on the lower and middle quantiles, policies to improve basic services and reduce poverty remain key components of long-term strategies. Conversely, in high-incidence areas, a more comprehensive policy approach is needed, incorporating climate data, population mobility, and clean and healthy living behaviors. Given its flexibility and adaptability, the model is recommended for implementation in real-time surveillance systems, enabling continuous monitoring and rapid response to emerging DHF trends at the local level.

Future studies are encouraged to expand this modeling approach by integrating spatial-temporal features, testing alternative regularization methods, and validating performance using real-time or panel datasets. This study confirms that statistical tools such as LASSO Quantile Regression are not only academically rigorous but also practically relevant in supporting evidence-based decision-making in public health, especially in the control of infectious diseases like DHF.

5. REFERENCES

[1] I. M. Sudarmaja, I. K. Swastika, L. P. E. Diarthini, I. P. D. Prasetya, and I. M. D. A. Wirawan, "Dengue virus transovarial transmission detection in Aedes aegypti from dengue hemorrhagic fever patients' residences in Denpasar, Bali," *Vet. World*, vol. 15, no. 4, pp. 1149–1153, 2022.

[2] A. Mercier *et al.*, "Impact of temperature on dengue and chikungunya transmission by the mosquito Aedes albopictus," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, 2022, [Online]. Available: https://doi.org/10.1038/s41598-022-10977-4

[3] K. Kesehatan, *Demam berdarah masih mengintai*, Edisi 165., no. April. 2024.

[4] M. W. Aditya, I. N. Sukajaya, and I. G. A. Gunadi, "Forecasting Jumlah Pasien DBD di BRSUD Kabupaten Tabanan Menggunakan Metode Regresi Linier," *Bali Med. J.*, vol. 10, no. 1, pp. 1–12, 2023.

[5] D. S. Susanti, P. D. Rahayu, and O. Soesanto, "Pemodelan Tingkat Kerawanan Demam Berdarah di Kabupaten Banjar dengan Metode Analisis Regresi Logistik yang terboboti Geografi," vol. 07, no. 01, pp. 52–64, 2019.

[6] R. Sriningsih, B. W. Otok, and Sutikno, "Determination of the best multivariate adaptive geographically weighted generalized Poisson regression splines model employing generalized cross-validation in dengue fever cases," *MethodsX*, vol. 10, no. February, p. 102174, 2023, [Online]. Available: https://doi.org/10.1016/j.mex.2023.102174

[7] T. Kim, B. Lieberman, G. Luta, and E. A. Peña, "Prediction intervals for Poisson-based regression models," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 14, no. 5, pp. 1–28, 2022.

[8] Annisa, A. Islamiyati, S. Sahriman, J. Massalesse, and U. Sari, "Truncated Spline Quantile Regression Model on Platelet Changes in Dengue Fever Patients Based on Body Temperature," *Commun. Math. Biol. Neurosci.*, pp. 1–11, 2024.

[9] I. Usman, A. Islamiyati, and E. T. Herdiani, "Quantile Regression Modeling With Group Least Absolute Shrinkage and Selection Operator Classification on Tuberculosis Data," *Commun. Math. Biol. Neurosci.*, vol. 2024, pp. 1–9, 2024.

[10] A. Anisa, A. Islamiyati, S. Sahriman, J. Massalesse, and B. Aprilia, "Model Regresi Kuantil Spline Orde Dua Dalam Menganalisis Perubahan Trombosit Pasien Demam Berdarah," *Jambura J. Math.*, vol. 5, no. 1, pp. 38–45, 2023.

[11] F. Rios-Avila and M. L. Maroto, "Moving Beyond Linear Regression: Implementing and Interpreting Quantile Regression Models With Fixed Effects," *Sociol. Methods Res.*, vol. 53, no. 2, pp. 639–682, 2024.

[12] B. Tantular, Y. Andriyana, and B. N. Ruchjana, "Quantile regression in varying coefficient model of upper respiratory tract infections in Bandung City," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021.

[13] J. P. Gygi, S. H. Kleinstein, and L. Guan, "Predictive overfitting in immunological applications: Pitfalls and solutions," *Hum. Vaccines Immunother.*, vol. 19, no. 2, 2023, [Online]. Available: https://doi.org/10.1080/21645515.2023.2251830

[14] M. S. H. Shaon, T. Karim, M. S. Shakil, and M. Z. Hasan, "A comparative study of machine learning models with LASSO and SHAP feature selection for breast cancer prediction," *Healthc. Anal.*, vol. 6, no. May, p. 100353, 2024, [Online]. Available: https://doi.org/10.1016/j.health.2024.100353

[15] C. Sun, B. Zhu, S. Zhu, L. Zhang, X. Du, and X. Tan, "Risk factors analysis of bone mineral density based on lasso and quantile regression in america during 2015–2018," *Int. J. Environ. Res. Public Health*, vol. 19, no. 1, 2022.

[16] Y. Sun and F. Lin, "Introduction and Some Recent Advances in Lp Quantile Regression," pp. 3827–3841, 2024.

[17] S. Wang, W. Cao, X. Hu, H. Zhong, and W. Sun, "A Selective Overview of Quantile Regression for Large-Scale Data," *Mathematics*, vol. 13, no. 5, pp. 4–6, 2025.

[18] W. N. A. Puteri, A. Islamiyati, and A. Anisa, "Penggunaan Regresi Kuantil Multivariat pada Perubahan Trombosit Pasien Demam Berdarah Dengue," *ESTIMASI J. Stat. Its Appl.*, vol. 1, no. 1, p. 1, 2020.

[19] G. Sottile and P. Frumento, "Robust estimation and regression with parametric quantile functions," *Comput. Stat. Data Anal.*, vol. 171, p. 107471, 2022, [Online]. Available: https://doi.org/10.1016/j.csda.2022.107471

[20] J. Chen *et al.*, "Study on the effect of occupational exposure on hypertension of steelworkers based on Lasso-Logistic regression model," *Public Health*, vol. 239, no. December 2024, pp. 15–21, 2025, [Online]. Available: https://doi.org/10.1016/j.puhe.2024.12.006

[21] M. Ouhourane, Y. Yang, A. L. Benedet, and K. Oualkacha, *Group penalized quantile regression*, vol. 31, no. 3. 2022.

[22] C. Ciner, B. Lucey, and L. Yarovaya, "Determinants of cryptocurrency returns: A LASSO quantile regression approach," *Financ. Res. Lett.*, vol. 49, no. May, p. 102990, 2022, [Online]. Available: https://doi.org/10.1016/j.frl.2022.102990

[23] J. Piironen, M. Paasiniemi, and A. Vehtari, "Projective inference in high-dimensional problems: Prediction and feature selection," *Electron. J. Stat.*, vol. 14, no. 1, pp. 2155–2197, 2020, doi: 10.1214/20-EJS1711.

[24] B. P. Statistik, *Catalog : 1101001*, vol. 53. 2025. [Online]. Available: https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html

[25] B. Dastjerdy, A. Saeidi, and S. Heidarzadeh, "Review of Applicable Outlier Detection Methods to Treat Geomechanical Data," *Geotechnics*, vol. 3, no. 2, pp. 375–396, 2023.
[26] J. Nyangon and R. Akintunde, "Principal component analysis of day-ahead electricity price forecasting in CAISO and its implications for highly integrated renewable energy markets," *Wiley Interdiscip. Rev. Energy Environ.*, vol. 13, no. 1, 2024.
[27] A. El Sheikh and M. R. Abonazel, "A Review of Penalized Regression and Machine Learning Methods in High-Dimensional Data," vol. 69, no. 1, pp. 250–261, 2025.
[28] C. Tian, N. Li, Y. Gao, and Y. Yan, "Analysis of the current status and influencing factors of oral frailty in elderly patients with type 2 diabetes mellitus in Taiyuan, China," *BMC Geriatr.*, vol. 25, no. 1, 2025.