JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 8, No. 2. Maret 2026 DOI: https://doi.org/10.34288/jri.v8i2.484
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

DENGUE FEVER CASE PREDICTION MODEL USING LINEAR REGRESSION WITH EXPLANATORY SEQUENTIAL MIXED METHODS APPROACH

Conchita Junita Chandra[1], Yoseph Thobias Pareira[2]

Informatics Engineering, Architecture
Nusa Nipa University
conchitachandra@gmail.com[1], yosephthobiaspareira@gmail.com[2]

Abstract

Dengue Hemorrhagic Fever (DHF) is an infectious disease in Indonesia, including in Sikka Regency, where the number of cases has increased over the past decade. Predicting the number of DHF cases is crucial to support disease prevention and control policies. This study aims to develop a predictive model for the number of dengue fever cases based on building area, population, and population density, moreover to explain other factors that influence the prediction results. The study uses an explanatory sequential mixed methods approach, and the prediction model is developed using simple linear regression and multiple linear regression. Quantitative data were obtained from the Health Office, the Sikka Regency Statistics Office, and Google Earth; while qualitative data were obtained through interviews with surveillance personnel from the Health Office and several community health centers in the study area, using a purposive sampling technique. The results show that the building area has a weak relationship with the number of DHF cases (R^2 = 0.10334 for Alok Timur sub-district and R^2=0.38055 for Waiblama). After adding the population and population density variables, the R^2 in Alok Timur increases to 0.46974; and R^2=0.41024 for Waiblama; however, the accuracy is still low. The interviews results show that community behavior is the dominant factors of DHF cases. This study indicates that predictive models based on physical environmental and population variables are unable to accurately depict the complexity of dengue fever case distribution. Therefore, the development of models that integrate community behavioral factors is necessary to provide more accurate predictions.

Keywords: Prediction, DHF, Building Area, Total Popoulation, Linear Regression

Abstrak

Demam Berdarah Dengue (DBD) merupakan salah satu penyakit menular yang menjadi masalah kesehatan masyarakat di Indonesia, termasuk di kabupaten Sikka-NTT yang cenderung meningkat dalam 10 tahun terakhir. Upaya memprediksi jumlah kasus DBD penting untuk mendukung kebijakan pencegahan dan pengendalian penyakit. Penelitian ini bertujuan untuk membangun model prediksi jumlah kasus DBD berdasarkan luas bangunan, jumlah penduduk, dan kepadatan penduduk, serta menjelaskan faktor lain yang berpengaruh terhadap hasil prediksi. Penelitian menggunakan pendekatan Explanatory Sequential Mixed Methods yang mengintegrasikan analisis kuantitatif dan kualitatif. Model prediksi dibangun secara bertahap menggunakan regresi linear sederhana dan regresi linear berganda. Data kuantitatif diperoleh dari Dinas Kesehatan, BPS kabupaten Sikka, dan Google Earth; sedangkan data kualitatif diperoleh melalui wawancara dengan surveilans dari Dinas Kesehatan dan beberapa puskesmas pada wilayah penelitian, dengan teknik purposive sampling. Hasil analisis menunjukkan bahwa luas bangunan memiliki hubungan yang lemah terhadap jumlah kasus DBD (nilai R^2=0,10334 untuk kecamatan Alok Timur dan R^2=0,38055 untuk kecamatan Waiblama). Setelah penambahan variabel jumlah penduduk dan kepadatan penduduk, nilai R^2 pada kecamatan Alok Timur meningkat menjadi 0,46974; dan R^2=0,41024 untuk kecamatan Waiblama; namun akurasi model masih tergolong rendah. Berdasarkan hasil wawancara, diketahui bahwa perilaku masyarakat, seperti kebiasaan menjaga kebersihan lingkungan dan kepatuhan terhadap program Pemberantasan Sarang Nyamuk, merupakan faktor dominan yang memengaruhi jumlah kasus DBD. Dapat disimpulkan bahwa model prediksi berbasis variabel lingkungan fisik dan kependudukan belum mampu menggambarkan kompleksitas penyebaran kasus DBD secara akurat. Oleh karena itu, diperlukan pengembangan model yang mengintegrasikan faktor perilaku masyarakat agar hasil prediksi lebih akurat untuk mendukung kebijakan pemerintah daerah terkait kasus DBD.

Kata kunci: Prediksi; DBD; Luas Area Bangunan; Jumlah Penduduk; Regresi Linear

INTRODUCTION

Dengue fever is an infectious disease that is a public health problem in tropical and subtropical climates (Chen & Moraga, 2025), including Indonesia (Mamenun et al., 2024; Ramadhani & Latif, 2021). Large cases of dengue fever occur in urban and semi-urban areas, especially during the rainy season, where environmental conditions support the breeding of Aedes Aegypti and Aedes Albopictus mosquitoes as intermediaries (vectors) (Kolimenakis et al., 2021; Puspita et al., 2023). DHF can cause death if not treated seriously (Ismail et al., 2024; Medagama, Dalugama, Meiyalakan, & Lakmali, 2020) and there is no effective warning system (Leung et al., 2023). World Health Organization (WHO) recorded the highest number of dengue cases in 2024, affecting more than 100 countries. More than 14.6 million cases and 12,000 deaths have been reported related to dengue fever. In fact, from January to July 2025, more than 4 million cases and 3,000 deaths have been reported to WHO from 97 countries, and it has spread to new regions including Europe and the Eastern Mediterranean (Organization, 2025). There is no specific treatment that significantly reduces the death rate from dengue fever (Tian et al., 2024), because often DHF does not show symptoms or only shows mild symptoms (Organization, 2025), and therefore the prevention and control of dengue fever depends on vector control (Koplewitz, Lu, Clemente, Buckee, & Santillana, 2022).

Sikka Regency, East Nusa Tenggara is one of the endemic areas for dengue fever and has been designated as an Extraordinary Event 5 times, which recurs every 3-4 years. From 2010 to 2019, there were 2,251 cases. The highest number of cases occurred in 2020 with 1,816 cases, with the largest age group being between 5 and 15 years old. In 2021, there were 183 cases, in 2022 with 466 cases, in 2023 with 822 cases, and in 2024 with 821 cases, and in early 2025, 50 cases were recorded, with 33 cases among children aged 5-15 years. Most cases occur in densely populated areas and urban areas (Adang, Marni, & Limbu, 2024; da Cunha, Samin, & Rahmawati, 2022; De Rosary, 2024; Pareira, Parera, & Hildegardis, 2023; Popi, 2025). According to data from the Sikka Regency Health Office, Alok Timur sub-district has the highest number of dengue fever cases, while Waiblama sub-district has the lowest number of cases. Therefore, these two locations were chosen for comparison in this study.

Based on data and previous research results in Sikka district, there is a significant relationship between the presence of mosquito larvae and dengue fever cases (Adang et al., 2024). The existence of larvae is influenced by environmental conditions (da Cunha et al., 2022; De Rosary, 2024; Popi, 2025), including the presence of buildings, which is associated with the number of mosquito breeding spots. Therefore, analyzing the relationship between building area and the number of dengue cases is important to understand the spread pattern of this disease and can develop more effective prevention strategies. Moreover, demographic factors such as population size and population density are also considered to have the potential to influence the number of dengue fever cases, because theoretically, areas with higher population size and density have a greater potential for disease transmission due to more intense interactions between individuals and limited living space.

This study aims to build and test the accuracy level of the prediction model for the number of dengue fever cases based on the building area, population and population density, using the linear regression method, which is a statistical method that can be used to determine the relationship between predictor variables (building area, population and population density) and response variables (number of dengue fever cases) (Roustaei, 2024). Linear regression is used because it is effective in analyzing the relationship between a dependent variable and one or more independent variables that are numerical in nature, in a simple and easy-to-interpret manner. In this case, all predictor variables (dependent variables) are numerical data, which theoretically relate to an increased risk of dengue fever (independent variables). This method is also effective for small data sizes, and can identify the variables that contribute most to variations in DHF cases. (Hope, 2020). In addition, residual analysis helps evaluate the accuracy of the model and reveals other factors that may influence the incidence of dengue fever (Kleden, Atti, & Sinu, 2025).

Comprehension of the correlation between the area of buildings, the number of residents and population density in an area with the number of DHF cases, the government through the Health Service and the Regional Development Planning, Research and Innovation Agency (BAPPERIDA), can use the available prediction models to map areas that are at high risk of increasing DHF cases. With the specific information, government can include more specific and optimal prevention programs and resource allocation such as DHF-related campaigns, medical personnel, medicines, and fogging equipment to locations that are predicted to

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.484
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

experience an increase in HDF cases. Additionally, the specific information is the basis for policy making related to spatial planning and development regulations.

RESEARCH METHODS

Types of research

This research uses the Explanatory Sequential Mixed Methods approach, namely a mixed research design that is carried out sequentially, starting with quantitative research, followed by qualitative research to explain the results of quantitative research (Draucker, Rawl, Vode, & Carter-Harris, 2020). The research focuses more on secondary data analysis to build a prediction model for the number of DHF cases based on physical environmental variables (building area) and population (total and density of population), nonetheless is supplemented by the results of interviews and observations related to supporting factors for DHF, such as community behavior, to support the conclusions. The research stages can be seen in Figure 1.

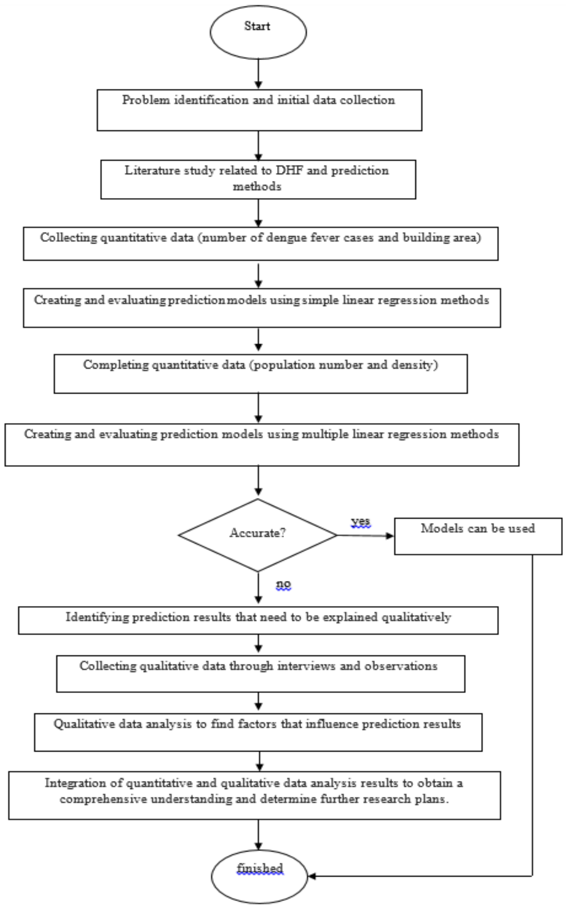

Figure 1. Research Flowchart

Time and Place of Research

This research was conducted in two areas in Sikka Regency: Alok Timur and Waiblama sub-districts. The study began with secondary data collection on the number of dengue fever cases through the Health Office and several community health centers in Sikka Regency. This was followed by site observations and interviews with health workers specializing in dengue fever surveillance at the Health Office and several community health centers from early July to August 2025. Data collection related to building area was also achieved using the Google Earth application from the end of July to the beginning of September 2025. Data processing, construction of a simple linear regression model using building area variables, model evaluation and analysis of the results were completed during September 2025. Secondary data collection was achieved in the form of population and population density data from the Central Statistics Agency in October 2025, followed by data processing, building models, evaluation and analysis of the results until mid-November 2025. Finnaly, the analysis was strengthened by interview results to obtain conclusions.

Research Target / Subject

The population in this study consists of all DHF case data in Alok Timur sub-district as the area with the highest number of cases and Waiblama sub-district with the lowest number of cases in Sikka district in 2016-2024. The subjects selected for interview are dengue fever surveillance officers at the Health Office and community health centers in the study area.

The sampling technique used was purposive sampling. (Nyimbili & Nyimbili, 2024), which was executed by selecting research areas based on the criteria of the largest number of cases and the area with the fewest number of cases over the last 10 years (extreme-case sampling). This region selection was decided to provide a contrasting picture of the influence of environmental and population factors on the number of dengue fever cases. For the interviews, the sample was selected using expert sampling (Firmansyah, 2022; Wahab, 2022), who are DHF surveillance officers.

Procedure

Overall, the research begins with a literature study, then determines the research area, data collection, data processing and analysis, evaluation, analysis of evaluation results, integrationn, and ends with formulating conclusions.

Specifically, the research was conducted in two main stages: quantitative and qualitative, following an Explanatory Sequential Mixed Methods design. The quantitative stage began with the collection of secondary data, including the number of dengue fever cases as the response variable and the building area as the predictor variable in the model design. The data was then analyzed to construct an initial prediction model, which was then evaluated and refined by adding population size and population density as predictor variables.

After obtaining quantitative results, the qualitative phase was directed by analyzing data from interviews with health workers to identify factors not included in the numerical data of the prediction results as important factors causing the low accuracy of the prediction model. The conclusion was reached that community behavior is one of the factors influencing the incidence of dengue fever. This integration of quantitative and qualitative results provides a comprehensive interpretation of the research findings.

Data, Instruments, and Data Collection Techniques

This study uses two types of data collected using an Explanatory Sequential Mixed Methods design. The quantitative data is consisted of secondary data from 2016 to 2024 for each research area, including:

1. The number of DHF cases per year as a response variable, which was obtained from the Sikka District Health Office and several community health centers in the research area;
2. Total building area as a predictor variable, which was obtained through the Google Earth application;
3. Population number and population density as predictor variables, wchich were obtained from the Central Statistics Agency of Sikka Regency.

This quantitative data is used to build a predictive model. Meanwhile, the qualitative data is consisted of semi-structured interviews with health workers (surveillance) at the Health Office and several community health centers in the study area, regarding the most frequently encountered factors in the field as causes of dengue fever. The results of the interviews, supplemented by the results of field observations in the form of photos of the presence of puddles and environmental cleanliness conditions, are used to explain the prediction results as quantitative findings, especially regarding variables that have a strong influence on the prediction model.

The tools used in collecting and analyzing quantitative data consist of:

1. Google Earth, for obtaining building area data in the research area;
2. Microsoft Excel, for managing quantitative data and manual modeling (as a comparison with models generated from Python software);
3. Python (Google Colabs), for building prediction models.

On the other hand, the tools used in collecting and analyzing qualitative data consist of: a semi-structured questionnaire as an interview guide and a mobile phone as a documentation tool.

The quantitative data collection techniques used in this study are included documentation and verification to ensure consistency between sources. Qualitative data collection techniques includes semi-structured interviews with expert informants and observations at the research site.

Data analysis technique

According to the Explanatory Sequential Mixed Methods approach, the analysis is carried out in three stages, namely:

1. Quantitative analysis, which aims to build a predictive model, with the following steps:
    a. Data pre-processing, which is finalized by:
        1) Checking the data inconsistencies, especially data on the number of dengue fever cases and population data;
        2) creating a derived variable (population density) which is calculated using the number of residents divided by the area of the residential area.
    b. The second step is developing a prediction model for the number of DHF cases based on building area by utilizing simple linear regression as shown with this following general equation:

$$Y = a + bX \qquad (1)$$

where:
Y = response variable (predicted value, in this case the number of dengue cases)
X = predictor variable (total building area in a sub-district)
a = intercept (Y value when X = 0)

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.484
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

b = regression coefficient (the amount of change in Y for every one unit change in X).

c. The next step is assessing model employs coefficient of determination value (R^2) and the average value of the prediction error (RMSE). R^2 displays how much of the proportion of variation in the response variable can be explained by the regression model based on its predictor variables.. The R^2 value ranges from 0-1. The larger the R^2 value (closer to 1) exposes the more accurate the prediction model (Chicco, Warrens, & Jurman, 2021). Meanwhile, RMSE measures how big the average error between the value predicted by the model and the actual value, in the same units (Chen & Moraga, 2025). The smaller the RMSE, the better the prediction model.
d. After assessing, the next step is enhancing predictor variables in the form of population and population density.
e. The next step is forming a prediction model by employing multiple linear regression as shown with this following general equation:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 \quad (2)$$

Where:
Y = number of dengue fever cases in a sub-district
X_1 = total building area in a sub-district
X_2 = number of popuplation in a sub-district area
X_3 = population density in a sub-district area
a = intercept (the value of Y when all X = 0)
b_1 = regression coefficient (average change in the number of DHF cases for every 1 m^2 increase in building area, with other predictors remaining constant)
b_2 = regression coefficient (average change in the number of DHF cases for each additional person, with other predictors constant)
b_3 = regression coefficient (average change in the number of DHF cases for every one person/km^2 increase in density, with other predictors constant).

f. The last step is re-evaluating the new model using R2 and RMSE.

The initial model showed inaccurate results. After adding predictor variables, the accuracy improved, but remained low, necessitating exploration of other factors.

2. Qualitative analysis, which aims to explore other factors not captured in the model that contribute to the low accuracy of the prediction model. The analysis was conducted thematically to identify the main factors influencing the number of dengue cases and the dominant themes that emerged related to community lifestyles.
3. The results of the qualitative and quantitative analysis are integrated to provide a comprehensive understanding and to support conclusions.

RESULTS AND DISCUSSION

The prediction model for the number of dengue fever cases was initially built using one predictor variable, namely the building area. Evaluation results indicate that:

1. The R^2 value for Alok Timur sub-district is very low, at 0.10334. This value indicates that building area only had a linear effect of 10.33% in predicting the number of dengue fever cases in Alok Timur sub-district. In contrast, the R^2 value in Waiblama sub-district, is 0.38055. Although theoretically, an R^2 value of >15% is acceptable and meaningful in the health sector due to the complexity of clinical phenomena (many factors influencing the prediction results), the RMSE value should also be considered (Gupta, Stead, & Ganti, 2024). The RMSE value in the Alok Timur sub-district area is 55.24, which means the average difference between the actual value and the predicted value is 55 cases. Conversely, the RMSE value in the Waiblama sub-district area is 2.24. Based on this value, the normalized Root Mean Square Error (nRMSE) value can be calculated using the equation:

$$nRMSE = \frac{RMSE}{Ymax - Ymin} \quad (3)$$

Where:
Ymax = the highest value of the range of values for the number of DHF cases (observe table 1)
Ymin = the lowest value of the range of values for the number of DHF cases (observe table 1)

P-ISSN: 2656-1743 | E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.484

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Thus, the nRMSE for East Alok is 31.75% and for Waiblama is 28%. According to forecasting interpretation guidelines, an nRMSE between 20-30% is acceptable. An nRMSE value >30% indicates a less accurate prediction of the number of dengue cases (Wang et al., 2021).

Table 1. Total Building Area and Number of Dengue Fever Victim Data

Year	Total Building Area		Number of Victim	
	Alok Timur	Waiblama	Alok Timur	Waiblama
2016	2329888.76	180132.83	70	1
2017	2383914.92	185793.29	25	1
2018	2409846.46	185793.29	33	0
2019	2455386.19	220676.17	97	2
2020	2500272.49	224559.42	199	8
2021	2547405.44	223976.84	36	0
2022	2581562.78	227237.95	74	7
2023	2610156.12	256798.47	142	5
2024	2451199.29	259868.99	161	5

2. The residual pattern in the graphs (Figure 2 and Figure 3) shows that the original data points are spread far from the regression line, which means that the linear line predicted by the model is not able to explain most of the variation in DHF cases (Lin & Liu, 2022).

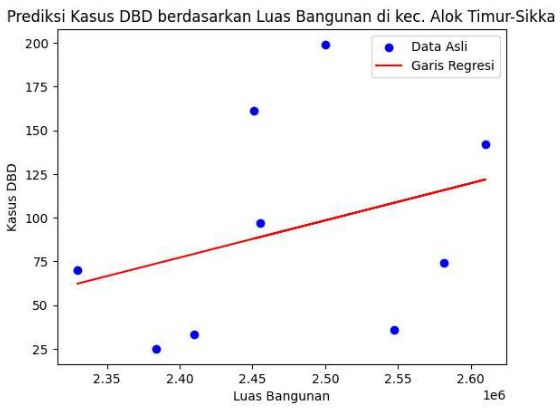

Figure 2. Comparison of Predicted Dengue Fever Cases Based on Building Area with Original Data in Alok Timur-Sikka

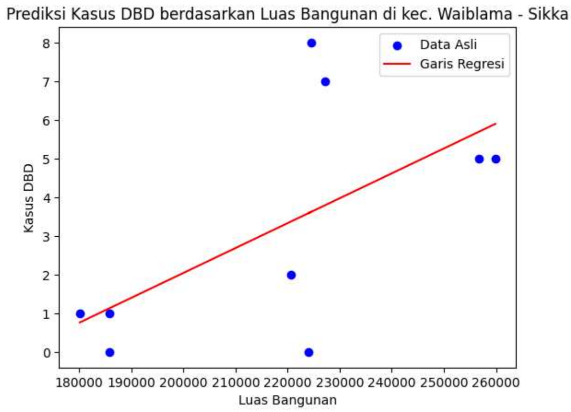

Figure 3. Comparison of Predicted Dengue Fever Cases Based on Building Area with Original Data in Waiblama-Sikka

Based on the evaluation analysis, it is concluded that building area alone is insufficient to accurately predict the number of dengue cases. Therefore, other variables, such as population size and population density, are added to improve model accuracy, resulting in several model variations with varying levels of accuracy. A comparison of model accuracy levels based on R^2 and RMSE values can be seen in Table 2.

Table 2. Comparison of R^2 and RMSE Values Based on Predictor Variables

Predictor	Alok Timur		Waiblama	
	R^2	RMSE	R^2	RMSE
Total building area	0,10334	55,24	0,38055	2,24
Total population	0,00154	58,30	0,34258	2,24
Population density	0,00153	58,30	0,34255	2,24
Total building area + Total population	0,11601	54,85	0,38588	2,24
Total building area + Population density	0,11597	54,85	0,38592	2,24

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.484
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Total building area + Total population + Population density	0,46974	42,48	0,41024	2,21

Regression analysis shows that building area, population and population density have a stronger influence on the number of dengue fever cases compared to other combinations of predictor variables, in all study locations. The nRMSE value for Alok Timur sub-district decreases from a poor rating (31.75%) to an acceptable rating (24.41%). Meanwhile, for Waiblama sub-district, it remains at the same rating, but with a percentage that also decreased from 28% to 27.63%. The prediction results based on the three predictor variables and the original data on the number of DHF cases for each sub-district can be seen in Figures 4 and 5. Although it shows an increase in the R2 value and a decrease in the RMSE and nRMSE values, the R^2 value is still relatively low and the RMSE is still quite high (as seen in Table 2), and the residual pattern is spread quite far from the regression line (as seen in Figure 6 and Figure 7) indicating that the three variables are only able to explain a small part of the variation in cases, so that the prediction accuracy is still limited. In addition, the predictor variables of population size and population density also have a multicollinearity relationship (Upendra, Abbaiah, & Balasiddamuni, 2023) (as seen from figure 8), thus, it is sufficient to choose one of them as a predictor. The models produced by each predictor can be seen in Table 2, both the R^2 and RMSE values are almost the same between the two models produced by the population number and population density predictors.

```
----------------------------------------
Koefisien (a): 0.00030462547585800723
Intercept (b): -36.9171405332163
----------------------------------------
----------------------------------------
   Luas Bangunan  Jumlah Penduduk  Kepadatan Penduduk  Prediksi Kasus DBD  \
0     2329888.76            33482              360.64           82.024861
1     2383914.92            33619              362.12           27.448562
2     2409846.46            33740              363.42           85.419551
3     2455386.19            32073              345.47           41.579608
4     2500272.49            32797              353.26          173.691653
5     2547405.44            32959              355.01          105.204209
6     2581562.78            33075              356.26          104.961915
7     2610156.12            33568              361.57          107.846176
8     2451199.29            34585              372.52          108.823466

   Kasus DBD Asli
0              70
1              25
2              33
3              97
4             199
5              36
6              74
7             142
8             161
----------------------------------------
----------------------------------------
R² Score: 0.46974309998751596
Mean Squared Error: 1805.2301573758346
----------------------------------------
```

Figure 4. Original Data and Predicted Results of the Number of Dengue Fever Cases Based on Building Area, Population, and Population Density in Alok Timur-Sikka

```
----------------------------------------
Koefisien (a): 0.00012073067992240911
Intercept (b): 4.918986125190309
----------------------------------------
----------------------------------------
   Luas Bangunan  Jumlah Penduduk  Kepadatan Penduduk  Prediksi Kasus DBD  \
0      180132.83             7277               50.41            0.789100
1      185793.29             7299               50.56            2.222789
2      185793.29             7317               50.69            0.326715
3      220676.17             8074               55.93            3.216348
4      224559.42             8074               55.93            3.685175
5      223976.84             8308               57.55            3.124189
6      227237.95             8363               57.93            3.668057
7      256798.47             8536               59.13            6.077744
8      259868.99             8929               61.85            5.889884

   Kasus DBD Asli
0               1
1               1
2               0
3               2
4               8
5               0
6               7
7               5
8               5
----------------------------------------
----------------------------------------
R² Score: 0.41024220293068736
Mean Squared Error: 4.951053111199168
----------------------------------------
```

Figure 5. Original Data and Predicted Results of the Number of Dengue Fever Cases Based on Building Area, Population, and Population Density in Waiblama-Sikka

P-ISSN: 2656-1743 | E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.484

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

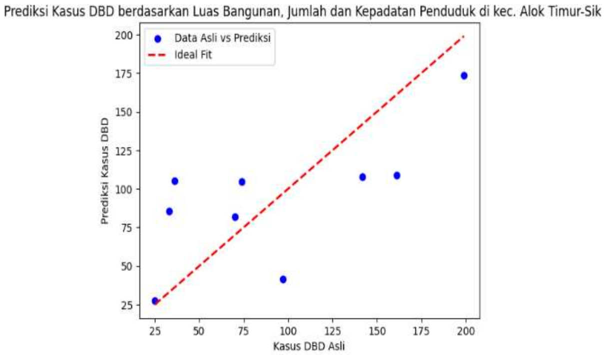

Figure 6. Comparison of Prediction Results of the Number of Dengue Fever Cases Based on Three Predictor Variables with Original Data from Alok Timur-Sikka

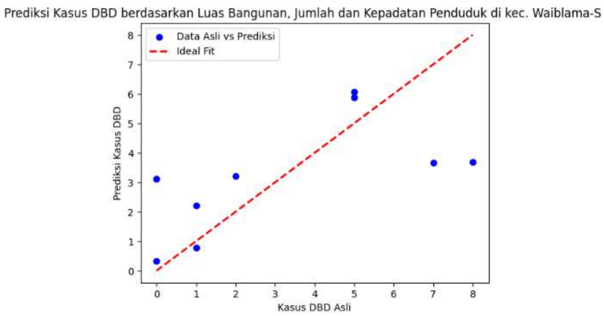

Figure 7. Comparison of Prediction Results of the Number of Dengue Fever Cases Based on Three Predictor Variables with Original Data from Waiblama-Sikka

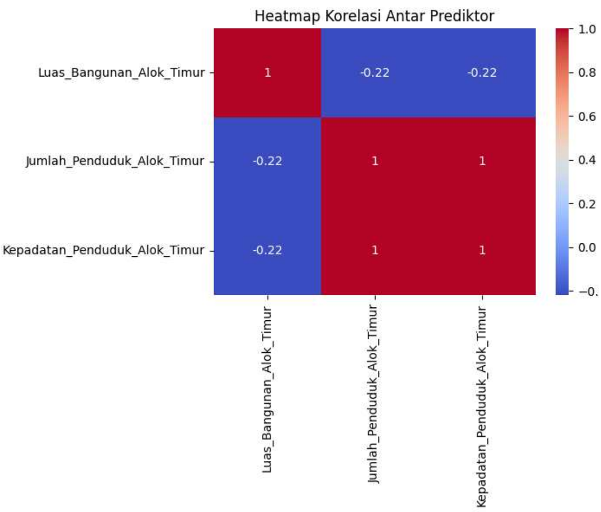

Figure 8. Correlation Between Predictors

The results of interviews with health workers at the Sikka District Health Office and several community health centers in the research area provides additional insight that is not apparent in the quantitative analysis: community behavior has a significant contribution to DHF cases. Those behaviors are:

1. allowing rubbish to pile up, especially plastic rubbish, which has the potential to collect rainwater;
2. The habit of storing water for a long time without covering it.

This is reinforced by the results of observations in areas with high DHF cases as seen in Figure 9.

Figure 9. Environmental Conditions in Alok Timur District

CONCLUSIONS AND SUGGESTIONS

Conclusion

The variations in prediction models based on physical and demographic environmental variables generated in this study are not optimal. Physical environmental factors (building area) and population density are influential, but not strong enough to accurately predict the number of dengue cases. This indicates that building area, population, and population density do not have a strong correlation with the number of dengue cases in Sikka Regency. Conversely, community behavior is an important factor causing dengue fever that is not captured in the prediction model. Therefore, the development of a model that integrates community behavioral factors is necessary.

Suggestion

Adding other predictor variables to the model is essential, especially those related to community behavior (from survey and interview results) with the intention to increase the accuracy of the model in predicting cases.

REFERENCES

Adang, T. E., Marni, & Limbu, R. (2024). Open

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.484
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Access. *Media Publikasi Promosi Kesehatan Indonesia The Indonesian Journal of Health Promotion*, *7*(6), 1627–1635. https://doi.org/https://doi.org/10.56338/mppki.v7i6.4719

Chen, X., & Moraga, P. (2025). Assessing dengue forecasting methods: a comparative study of statistical models and machine learning techniques in Rio de Janeiro, Brazil. *Tropical Medicine and Health*, *53*(1), 52. https://doi.org/10.1186/s41182-025-00723-7

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

da Cunha, M. F. C., Samin, M., & Rahmawati, A. (2022). HUBUNGAN PERILAKUWARGA DI RUMAH DENGAN KONDISI LINGKUNGAN TEMPAT TINGGAL SEBAGAI FAKTOR PENYEBAB DEMAM BERDARAH DENGUE DI KELURAHAN KOTA UNENG KECAMATAN ALOK KABUPATEN SIKKA. *Jurnal Geografi*, *18*(1), 126–146. Retrieved from https://ejurnal.undana.ac.id/index.php/jgeo/article/view/7526

De Rosary, E. (2024). NTT Endemik Demam Berdarah, Bagaimana Penanganannya? Retrieved March 26, 2025, from https://mongabay.co.id/2024/04/13/ntt-endemik-demam-berdarah-bagaimana-penanganannya/

Draucker, Claire Burke, Rawl, Susan M, Vode, Emilee, & Carter-Harris, Lisa. (2020). Integration Through Connecting in Explanatory Sequential Mixed Method Studies. *Western Journal of Nursing Research*, *42*(12), 1137–1147. https://doi.org/10.1177/0193945920914647

Firmansyah, D. D. (2022). Teknik pengambilan sampel umum dalam metodologi penelitian: Literature review. *Jurnal Ilmiah Pendidikan Holistik (JIPH)*, *1*(2), 85–114.

Gupta, A., Stead, T. S., & Ganti, L. (2024). Determining a meaningful R-squared value in clinical medicine. *Academic Medicine \& Surgery*.

Hope, T. M. H. (2020). *Chapter 4 - Linear regression* (A. Mechelli & S. B. T.-M. L. Vieira, Eds.). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-815739-8.00004-3

Ismail, N. A., Haikhah, J., Fauziah, E. T., Prabasunu, M. A., Rahma, F. H., Siswanto, S., & Widyasari, V. (2024). Investigation of risk factors for household-based dengue virus infection in Borobudur Subdistrict, Magelang, Indonesia. *Germs*, *14*(3), 277–286. https://doi.org/10.18683/germs.2024.1438

Kleden, M. A., Atti, A., & Sinu, E. B. (2025). Spatial Modeling of Climate Change Effects on Dengue Fever Incidence: A Case Study in East Nusa Tenggara. *Mathematical Modelling of Engineering Problems*, *12*(1), 309–328. https://doi.org/10.18280/mmep.120131

Kolimenakis, A., Heinz, S., Wilson, M. L., Winkler, V., Yakob, L., Michaelakis, A., … Horstick, O. (2021). The role of urbanisation in the spread of Aedes mosquitoes and the diseases they transmit—A systematic review. *PLOS Neglected Tropical Diseases*, *15*(9), e0009631. Retrieved from https://doi.org/10.1371/journal.pntd.0009631

Koplewitz, G., Lu, F., Clemente, L., Buckee, C., & Santillana, M. (2022). Predicting dengue incidence leveraging internet-based data sources. A case study in 20 cities in Brazil. *PLOS Neglected Tropical Diseases*, *16*(1), e0010071. https://doi.org/10.1371/journal.pntd.0010071

Leung, X. Y., Islam, R. M., Adhami, M., Ilic, D., McDonald, L., Palawaththa, S., … Karim, N. (2023). A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Neglected Tropical Diseases*, *17*(2), e0010631. https://doi.org/10.1371/journal.pntd.0010631

Lin, Z., & Liu, D. (2022). *Model diagnostics of discrete data regression: a unifying framework using functional residuals*. Retrieved from https://arxiv.org/abs/2207.04299

Mamenun, Koesmaryono, Y., Sopaheluwakan, A., Hidayati, R., Dasanto, B. D., & Aryati, R. (2024). Spatiotemporal Characterization of Dengue Incidence and Its Correlation to Climate Parameters in Indonesia. *Insects*, Vol. 15. https://doi.org/10.3390/insects15050366

Medagama, A., Dalugama, C., Meiyalakan, G., & Lakmali, D. (2020). Risk Factors Associated with Fatal Dengue Hemorrhagic Fever in Adults: A Case Control Study. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *2020*(1), 1042976. https://doi.org/https://doi.org/10.1155/20

20/1042976

Nyimbili, F., & Nyimbili, L. (2024). Types of purposive sampling techniques with their examples and application in qualitative research studies. *British Journal of Multidisciplinary and Advanced Studies: English Lang., Teaching, Literature, Linguistics & Communication*, *5*(1), 90–99. https://doi.org/https://doi.org/10.37745/bjmas.2022.0419

Organization, W. H. (2025). Dengue. Retrieved November 6, 2025, from https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

Pareira, Y. T., Parera, Y. P. P., & Hildegardis, C. (2023). PENGARUH SETTING FISIK LINGKUNGAN TERHADAP KEJADIAN DBD (DEMAM BERDARAH DENGUE) BERDASARKAN KARAKTERISTIK TERMAL DI KABUPATEN SIKKA, NUSA TENGGARA TIMUR. *Jambura Journal of Architecture*, *5*(1). https://doi.org/https://doi.org/10.37905/jjoa.v5i1.19163

Popi, P. (2025). Dinkes Sikka Catat 50 Kasus DBD di Awal 2025, Anak-anak Paling Banyak Terinfeksi. Retrieved March 26, 2025, from https://ekorantt.com/2025/02/11/dinkes-sikka-catat-50-kasus-dbd-di-awal-2025-anak-anak-paling-banyak-terinfeksi/

Puspita, J. W., Farida, Fakhruddin, M., Nuraini, N., Fauzi, R., Indratno, S. W., & Soewono, E. (2023). Modeling and descriptive analysis of dengue cases in Palu City, Indonesia. *Physica A: Statistical Mechanics and Its Applications*, *625*, 129019. https://doi.org/https://doi.org/10.1016/j.physa.2023.129019

Ramadhani, S. N., & Latif, M. T. (2021). IMPACT OF CLIMATE CHANGE ON DENGUE HEMORRHAGIC FEVER (DHF) IN TROPICAL COUNTRIES : *Jurnal Kesehatan Lingkungan*, *13*(4), 219–226. https://doi.org/10.20473/jkl.v13i4.2021.219-226

Roustaei, N. (2024). Application and interpretation of linear-regression analysis. *Medical Hypothesis, Discovery & Innovation Ophthalmology Journal*, *13*(3), 151–159. https://doi.org/10.51329/mehdiophthal1506

Tian, N., Zheng, J.-X., Li, L.-H., Xue, J.-B., Xia, S., Lv, S., & Zhou, X.-N. (2024). Precision Prediction for Dengue Fever in Singapore: A Machine Learning Approach Incorporating Meteorological Data. *Tropical Medicine and Infectious Disease*, Vol. 9, p. 72. https://doi.org/10.3390/tropicalmed9040072

Upendra, S., Abbaiah, R., & Balasiddamuni, P. (2023). Multicollinearity in Multiple Linear Regression : Detection , Consequences , and Remedies. *International Journal for Research in Applied Science and Engineering Technology*, (September). https://doi.org/https://doi.org/10.22214/ijraset.2023.55786

Wahab, A. J. (2022). Sampling dalam Penelitian Kesehatan. *Jurnal Pendidikan Dan Teknologi Kesehatan*, *5*(1), 42–49.

Wang, P., Ma, Y., Tang, J., Wu, D., Chen, H., Jin, Z., & Huo, Z. (2021). Spring Frost Damage to Tea Plants Can Be Identified with Daily Minimum Air Temperatures Estimated by MODIS Land Surface Temperature Products. *Remote Sensing*, Vol. 13, p. 1177. https://doi.org/10.3390/rs13061177