International Journal of Computer Technology and Science

E-ISSN: 3048-1961
P-ISSN: 3048-1899

Research Article

Advanced Machine Learning for Comprehensive Mapping and Risk Analysis of Dengue Fever in Purwokerto to Support Public Health Preparedness

Rosa Ratri Kusuma Hariningsih[1*], Diwahana Mutiara Candrasari[2], Endang Setyawati[3], Syamsu Wahidin[4], Jevon Nataniel Putra[5]

[1] Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia 1; e-mail : rosaratri23@gmail.com
[2] Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia 2; e-mail : candrasari5860@stikomyos.ac.id
[3] Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia 3; e-mail : endang.setiawati@stikomyos.ac.id
[4] Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia 4; e-mail : syamsuwahidin@gmail.com
[5] Sekolah Tinggi Ilmu Komputer Yos Sudarso, Indonesia 5; e-mail : yevonthemythhunt1@gmail.com
* Corresponding Author: Rosa Ratri Kusuma Hariningsih

Abstract: Dengue Fever (DF) continues to be a major public health threat in Indonesia, especially in urban areas with high population density, such as Purwokerto City. This study aims to develop a predictive model to identify high-risk areas for DF outbreaks by integrating Machine Learning (ML) algorithms and Geographic Information Systems (GIS). The research utilizes historical dengue case data, meteorological parameters (rainfall, temperature, humidity), and population density as predictive variables. Three ML classification algorithms—Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM)—were implemented to develop risk prediction models. Extensive data preprocessing, feature selection, and spatial integration were applied to ensure model robustness. The results show that the SVM model outperformed other methods, achieving the highest accuracy, precision, recall, and F1-score in classifying dengue risk zones. Risk maps generated through GIS visualization successfully identify priority areas for targeted interventions. The novelty of this research lies in the combination of local epidemiological data, multi-algorithm comparison, and geospatial mapping to improve early warning systems for DF in Purwokerto. This integrated approach is expected to support more effective prevention strategies and enhance public health preparedness.

Keywords: Dengue Fever; GIS; Machine Learning; Public Health Preparedness; Risk Mapping.

Received: June 07, 2025
Revised: June 21, 2025
Accepted: July 05, 2025
Published: July 07, 2025
Curr. Ver.: July 07, 2025

1. Introduction

Dengue Fever (DF) remains a significant global public health concern, particularly in tropical countries such as Indonesia [1]. The disease is caused by dengue virus infection, transmitted primarily through the bites of Aedes aegypti and Aedes albopictus mosquitoes [2]. Environmental factors, including high rainfall, warm temperatures, elevated humidity, and high population density, significantly influence the dynamics of DF transmission in affected regions [3].

In recent years, DF outbreaks have become increasingly difficult to control due to the complexity of contributing factors, including climate change, rapid urbanization, and high levels of human mobility [4]. This situation is also evident in Banyumas Regency, Central Java Province. Data from the Banyumas Health Office reported a significant surge in DF cases in early 2024, with a total of 135 cases recorded within the period of January to February alone [5]. This increase is partly attributed to extreme fluctuations in rainfall, which create ideal conditions for the proliferation of Aedes mosquito populations—the primary vectors of the disease [6].

Purwokerto City, as the administrative and economic center of Banyumas Regency, exhibits a relatively high risk of DF transmission. The city consists of four main sub-districts: Purwokerto Timur (East), Purwokerto Barat (West), Purwokerto Selatan (South), and Purwokerto Utara (North), which consistently report higher DF case numbers compared to surrounding areas. Therefore, innovative efforts are required to enhance preparedness and improve the effectiveness of DF prevention strategies.

One promising approach involves the application of predictive modeling based on Machine Learning (ML) techniques [7]. Various methods such as Naïve Bayes Classifier, Binary Logistic Regression, and Support Vector Machine (SVM) have been widely adopted to predict and classify the occurrence of specific phenomena by analyzing historical data patterns [8]. These models offer advantages in terms of predictive accuracy, computational efficiency, and the ability to handle datasets with multiple interrelated variables, including climatic factors such as temperature, humidity, rainfall, and DF case distribution [9].

In addition, the integration of Geographic Information Systems (GIS) technology enables the spatial visualization of prediction results in the form of risk zone maps [10]. The availability of data-driven predictive systems combined with spatial risk maps is expected to optimize community preparedness and facilitate more targeted intervention efforts in high-risk areas [11].

This study aims to develop a predictive model for DF high-risk areas in Purwokerto City by integrating Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM) algorithms with spatial analysis using GIS technology [12]. Model performance evaluation is conducted to determine the most optimal method for DF risk classification [13]. Furthermore, the prediction results are visualized through spatial risk zone maps for DF distribution [14]. Through this integrated approach, the study is expected to contribute to improving the effectiveness of mitigation strategies, early detection, and evidence-based decision-making for controlling DF transmission, particularly in urban areas with high population density, such as Purwokerto City.

2. Literature Review

This section presents the theoretical foundation, state-of-the-art research, and existing gaps in the field of dengue fever (DF) prediction and spatial mapping. Several relevant studies are discussed, focusing on the use of Machine Learning (ML), Geographic Information Systems (GIS), and their integration for disease risk prediction and early warning systems.

2.1 Dengue Fever, Environmental Factors, and Risk Mapping

Dengue Fever (DF) remains a significant public health issue in tropical regions, with Indonesia experiencing fluctuating outbreaks driven by environmental and demographic factors. Previous studies have highlighted the role of rainfall, temperature, humidity, and population density as key contributors to DF transmission patterns. For instance, increased rainfall creates ideal breeding grounds for Aedes mosquitoes, while optimal temperatures between 26°C and 29°C accelerate their life cycle.

Spatial risk mapping has been recognized as an effective approach to visualize high-risk areas and guide targeted interventions. GIS tools, such as heatmaps and clustering techniques, are widely used to analyze disease distribution patterns and environmental determinants.

However, conventional statistical models often struggle to capture complex, non-linear interactions among risk factors, limiting their predictive accuracy. This limitation has driven the adoption of Machine Learning algorithms for DF risk prediction.

2.2 Machine Learning Applications for Dengue Risk Prediction

Machine Learning offers powerful tools to handle complex epidemiological data, providing higher predictive performance compared to traditional models. Several notable studies have demonstrated the application of ML in DF risk assessment:

a. Rattanavipanon et al. (2020) utilized Machine Learning and GIS to map Aedes aegypti distribution in Thailand. While effective in predicting mosquito abundance, the study did not provide direct spatial risk maps for DF cases.
b. Rahman et al. (2020) applied the Naïve Bayes algorithm to predict DF outbreak risk in tropical regions, showing promising results. However, the research only used a single algorithm without comparing alternative models or including spatial visualization.
c. Santos et al. (2021) demonstrated the use of ML for dengue risk assessment in Brazil, integrating environmental and epidemiological data. Nevertheless, the study focused

on regional scales and lacked emphasis on localized, micro-environmental variations within cities.

d. Hasan et al. (2021) implemented K-Means clustering to analyze DF risk areas in Indonesia, providing national-level insights but neglecting urban-level variations, such as district-specific population density and sanitation differences.
e. Wang et al. (2022) applied ML algorithms to predict the spread of DF and other infectious diseases, emphasizing the potential of these methods for outbreak prevention.

These studies collectively confirm the advantages of ML for disease risk prediction. However, they also reveal research gaps, particularly regarding:

a. Limited integration of local epidemiological data with spatial risk mapping at the district or sub-district level.
b. Insufficient model comparisons to identify the most effective ML algorithm for DF early warning systems.
c. Lack of localized studies that account for micro-environmental differences, especially in densely populated urban areas like Purwokerto, Indonesia.

3. Proposed Method

This study employs a quantitative approach with a predictive research design. The primary objective is to develop a risk prediction model for Dengue Fever (DF) outbreaks in Purwokerto by utilizing historical case data and relevant environmental variables. The research process consists of four main stages: data collection, predictive model development, model performance evaluation, and spatial visualization of prediction results.

3.1 Data Collection and Preprocessing

The data used in this study were obtained from two main sources:

a. Dengue Fever case data for the period 2022–2024, provided by the Banyumas Regency Health Office.
b. Meteorological parameters, including air temperature, humidity, and rainfall, provided by the Meteorology, Climatology, and Geophysics Agency (BMKG) of Indonesia.

The initial preprocessing stage involved the following steps:

a. Data cleaning to remove missing values, duplicates, and outliers.
b. Spatial mapping of the data based on geographic coordinates, aligning the data with the administrative boundaries of Purwokerto.
c. Transformation of the target variable into binary categories, where 1 represents high risk (more than 5 DF cases) and 0 represents low risk (5 or fewer DF cases).
d. Selection of key independent variables for modeling, including air temperature (°C), humidity (%), rainfall (mm), and the month of occurrence.
e. Feature standardization using the StandardScaler method, except for the Naïve Bayes model, which does not require normalization due to its probabilistic nature.

3.2 Predictive Model Development

The preprocessed dataset was divided into training (70%) and testing (30%) subsets using stratified sampling to maintain the proportional distribution of target classes. Three machine learning algorithms were applied in this study:

a. Support Vector Machine (SVM)

SVM constructs an optimal hyperplane to separate the two classes. The parameter class_weight='balanced' was applied to address class imbalance issues [15]. Hyperparameter optimization for SVM was performed using Grid Search Cross-Validation (CV) by tuning the C, kernel, and gamma parameters [16].

Support Vector Machine (SVM) is an algorithm designed for more complex classification tasks and is highly effective in handling high-dimensional data [9]. SVM operates by finding the optimal hyperplane that separates two distinct classes of data [25]. The decision boundary can be mathematically defined as:

$$f(x) = w^T x + b = 0$$

$f(x)$ is the decision function,
w is the weight vector,
x represents the feature vector,
b is the bias or intercept term.

The primary objective of SVM is to maximize the margin between the two classes, ensuring better generalization and robustness of the model. The optimization is performed under the following constraints:

$$y_i \epsilon \{-1,1\}$$

y_i denotes the class label of each training data point,
x_i represents the training data samples.

By maximizing the margin while satisfying these constraints, SVM aims to create a decision boundary that best separates the classes with minimal classification error.

b. Logistic Regression

This algorithm served as the baseline binary classification model, providing probabilistic output [17]. Hyperparameters such as C, solver, and max_iter were optimized using Grid Search CV, with class balancing applied to handle data imbalance [18].

Binary Logistic Regression is an algorithm used to predict the probability of an outcome with two possible categories (for example, whether a person is infected or not) [24]. This model is widely applied in binary classification problems, such as email spam detection (spam/not spam) or medical diagnosis (positive/negative) [18].

The mathematical formulation of Binary Logistic Regression is expressed as:

$$\boldsymbol{P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}}}$$

$P(Y = 1|X)$ represents the probability that the outcome Y = 1 given the predictor variables X.
β_0 is the intercept (constant term)
β_i are the regression coefficients corresponding to the independent variables X_i
X_i denotes the independent variables
e is Euler's number (approximately 2.718)\

The parameters β in logistic regression are estimated using the Maximum Likelihood Estimation (MLE) method, which seeks to find the parameter values that maximize the likelihood function:

$$L(\beta) = \prod_{i=1}^{m} P(Y_i|X_i)$$

Where *m* is the total number of samples in the dataset. MLE ensures that the estimated model parameters provide the highest probability of observing the given data.

c. Multinomial Naïve Bayes

A probabilistic model based on Bayes' Theorem, which assumes independence among features. Due to this assumption, standardization was not required for this model [19].

Naïve Bayes Classifier is an algorithm used to predict the probability of a classification based on known conditions [22]. Naïve Bayes is a classification algorithm that applies Bayes' Theorem with the fundamental assumption that all features within the dataset are conditionally independent from one another [23]. The Bayes' Theorem can be formulated as follows:

$$P(C_k|X) = \frac{P(X|C_k)P(C_k)}{P(X)}$$

$P(C_k|X)$ represents the posterior probability of class C_k iven the features X (posterior probability)

$P(X|C_k)$ is the likelihood, i.e., the probability of observing the features X given class C_k (likelihood)
$P(C_k)$ denotes the prior probability of class C_k (prior probability)
$P(X)$ is the marginal probability of the features X(also known as the evidence).
Due to the assumption of conditional independence between features, the likelihood term can be decomposed as follows:

$$P(X|C_k) = P(x_1|C_k)P(x_2|C_k) \dots P(x_n|C_k)$$

Thus, the Naïve Bayes formula simplifies to:

$$P(C_k|X) \propto P(C_k)\prod_{i=1}^{n} P(x_i|C_k)$$

3.3 Model Performance Evaluation

Model performance was assessed using five key evaluation metrics:

a. Accuracy: The proportion of correct predictions over the total test data.
b. Precision: The model's ability to correctly predict positive (high-risk) cases.
c. Recall (Sensitivity): The model's ability to identify all actual positive (high-risk) cases.
d. F1-Score: The harmonic mean of precision and recall.
e. Confusion Matrix: A table summarizing correct and incorrect predictions for each class.

The evaluation results were presented in summary tables, heatmaps of the confusion matrices, and bar charts comparing the performance of each model. Based on these evaluations, the Support Vector Machine (SVM) demonstrated the best overall performance and was selected as the primary model for risk mapping.

3.4 Risk Mapping Using GIS

The best-performing SVM model was utilized to predict the DF risk level for each data entry based on environmental and temporal variables [20]. The prediction results were integrated with the administrative spatial data of Purwokerto.

The predicted data, labeled with risk classes (0 = low risk, 1 = high risk), were exported in GeoJSON format, containing both the predictive attributes and spatial geometry of the regions. These files were imported into QGIS software and visualized thematically using graduated color schemes based on the predicted risk levels.

This visualization process produced spatial risk distribution maps for DF, with distinct color representations differentiating areas based on risk severity. These maps serve as visual decision-support tools to facilitate prioritization of preventive interventions in areas classified as high risk for DF outbreaks.

3.5 Model Preservation

To ensure reusability and efficiency for long-term application, the optimized SVM model and the corresponding feature standardization process were saved in .pkl format using the Joblib library [21]. This allows for future deployment of the model on new data without the need for retraining and facilitates integration with digital monitoring platforms for real-time DF risk surveillance.

4. Results and Discussion

This study aims to develop a predictive model for the spatial distribution of Dengue Fever (DF) cases in Purwokerto City using machine learning techniques and geospatial mapping. The research results are explained in detail, including data description, exploratory analysis, model evaluation, and spatial risk mapping based on the prediction outcomes.

4.1 Data Description and Exploratory Analysis

The dataset used in this study consists of several variables that influence the spread of DF, including the number of DF cases, rainfall, air temperature, and humidity. The data were collected for the period 2022 to 2024 from two main sources: the Banyumas Regency Health Office and the Meteorology, Climatology, and Geophysics Agency (BMKG) of Indonesia. The research focuses on four sub-districts in Purwokerto City, identified as having high population density: Purwokerto Timur (East), Purwokerto Barat (West), Purwokerto Selatan (South), and Purwokerto Utara (North).

The initial exploratory analysis revealed an increasing trend in DF cases, which was positively correlated with both rainfall and humidity levels. Data visualization showed that the peak of DF cases occurred during the early months of the year (January to March), coinciding with the period of highest rainfall intensity. Air temperature in the region generally ranged between 26°C and 29°C, which is considered the optimal range for the breeding and development of Aedes aegypti mosquitoes. Among the four sub-districts, Purwokerto Timur and Purwokerto Utara reported the highest incidence rates of DF compared to other areas.

4.2 Machine Learning Modeling

This study applied three Machine Learning (ML) algorithms, namely Binary Logistic Regression, Naïve Bayes Classifier, and Support Vector Machine (SVM). These models were developed to classify areas into high-risk and low-risk zones for Dengue Fever (DF) transmission, based on environmental and population-related features.

a. Model Implementation

The processed dataset was analyzed using k-fold cross-validation (k = 5). Each model was trained to classify sub-districts into high-risk or low-risk categories for DF. The dataset was divided into 70% training data and 30% testing data.

Model	Accuracy	Precision (0)	Recall (0)	F1-Score (0)	Precision (1)	Recall (1)	F1-Score (1)	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score
SVM Tuning	0.82	0.97	0.81	0.88	0.46	0.86	0.60	0.71	0.83	0.74
Logistic Regression	0.48	0.85	0.46	0.60	0.17	0.57	0.26	0.51	0.52	0.43
Multinomial Naive Bayes	0.55	0.90	0.51	0.66	0.22	0.71	0.33	0.56	0.61	0.49

Figure 1 Three Model Prediction Results

b. Model Evaluation

The models were evaluated using accuracy, precision, recall, and F1-Score metrics for both classes (class 0 = low risk, class 1 = high risk). The evaluation was conducted using k-fold cross-validation and a 70:30 train-test data split. The summary of model performance is as follows:

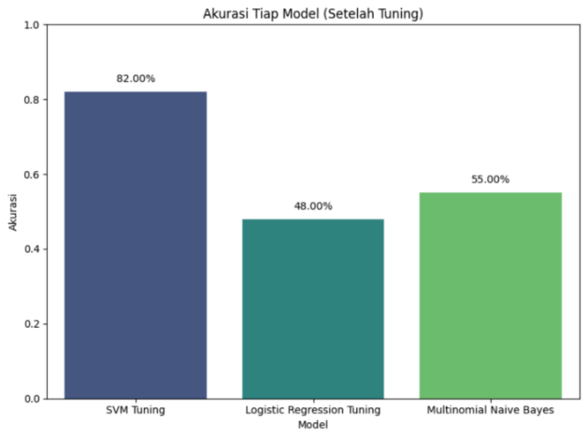

Figure 2 Evaluation of Three Machine Learning Models

Based on the evaluation results, the tuned SVM model demonstrated the best performance among the three algorithms. The SVM achieved an accuracy of 82% and a recall of 0.86 for the high-risk class (class 1), indicating the model's strong ability to detect high-risk areas effectively. The macro-average precision (0.71), recall (0.83), and F1-Score (0.74) further confirmed the superior performance of the SVM model compared to Logistic Regression and Naïve Bayes, which tended to underperform, particularly in positive class classification.

These results indicate that the tuned SVM model excels not only in terms of overall accuracy but also in maintaining a good balance between precision and recall for both classes, especially in identifying high-risk areas with a recall of 0.86. The higher macro-average across all key evaluation metrics further solidifies the SVM model as the best-performing approach for DF risk mapping in this study.

4.3 Dengue Fever Risk Prediction Mapping Using SVM

The trained and evaluated SVM was then deployed to classify each area (based on environmental and population data) into high-risk or low-risk zones. The classification results were integrated with spatial data in a Geographic Information System (GIS) using QGIS software.

The classification outcomes were visualized as thematic maps (heatmaps), illustrating the spatial distribution of DF risk across the four sub-districts in Purwokerto City. Red zones indicate high-risk areas, while green zones represent low-risk areas. The results are summarized as follows:

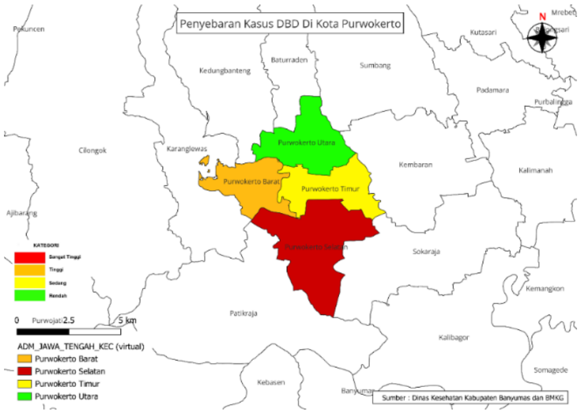

Figure 3 Dengue Fever Case Risk Zoning Map in Purwokerto

a. Purwokerto Selatan (South) and Purwokerto Barat (West) were identified as high-risk areas consistently throughout the observation period (2022–2024), particularly in densely populated neighborhoods with extreme rainfall patterns.
b. Purwokerto Timur (East) was classified as a moderate-risk area, with an increasing risk trend observed in 2024.
c. Purwokerto Utara (North) was generally categorized as a low-risk area; however, it requires continuous monitoring, especially in the event of rising temperatures and humidity levels.

The SVM-based risk maps provide concrete insights for the local health authorities and government to determine priority areas for interventions such as fogging operations, community education campaigns, and enhanced public health preparedness.

4.4 Implications

Based on the analysis results, the following findings were obtained:

a. The Support Vector Machine (SVM) consistently outperformed Logistic Regression and Naïve Bayes models in terms of accuracy and spatial classification capability.
b. Environmental factors such as temperature, rainfall, humidity, and population density exhibited strong correlations with DF transmission risk, with optimal temperature ranges and high rainfall identified as major drivers of mosquito population growth.
c. The integration of SVM predictions with GIS mapping effectively identified high-risk areas, which are often difficult to detect using conventional statistical approaches.
d. The application of Machine Learning and GIS in this study resulted in a data-driven, localized early warning system that can be integrated into public health mitigation policies by relevant stakeholders.

5. Comparison

The results of this study demonstrate that the integration of Machine Learning and spatial data provides accurate and practical predictions for area-based disease control. The application of three classification models, namely Logistic Regression, Naïve Bayes, and Support Vector Machine (SVM), revealed significant differences in performance, with SVM consistently delivering the best results in classifying Dengue Fever (DF) risk zones in Purwokerto City.

The SVM model proved superior in terms of both overall accuracy and its ability to detect high-risk areas, as indicated by a high recall score for the positive class (class 1). This finding suggests that the SVM model is highly sensitive to environmental patterns that trigger DF outbreaks. These results are consistent with previous studies indicating that SVM is effective in handling high-dimensional datasets and capturing non-linear patterns in public health phenomena.

Furthermore, the success of spatial DF risk mapping highlights that integrating classification results with Geographic Information Systems (GIS) can serve as a highly effective decision-support tool. The risk visualization in the form of heatmaps enables relevant authorities to quickly identify priority intervention areas based on reliable, evidence-driven insights. Such a strategy is crucial for managing infectious diseases like DF, which require rapid, localized responses.

Importantly, this study also confirms that environmental factors—including rainfall, humidity, air temperature, and population density—play a significant role in DF transmission. These findings strengthen the existing literature, which emphasizes ecological and climatic factors as key determinants in the life cycle of Aedes aegypti mosquitoes and the spread of the dengue virus.

Nevertheless, several limitations were identified in this research, including the limited spatial coverage, which only encompassed four sub-districts, and the absence of real-time data or direct mosquito vector data. Future work should expand the model's geographical coverage and integrate it with a web-based monitoring dashboard and early warning system for the community.

Overall, considering the results and analyses, the Machine Learning approach based on SVM combined with GIS mapping has been proven to make a significant contribution to early detection efforts and evidence-based policy-making for DF outbreak mitigation in Purwokerto City.

6. Conclusions

This study successfully demonstrated that the integration of Machine Learning (ML) algorithms and Geographic Information Systems (GIS) is an effective approach for predicting and mapping Dengue Fever (DF) risk areas in Purwokerto City. The predictive models developed using Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes algorithms showed that SVM delivered the best performance, achieving an overall accuracy of 82% and a recall score of 0.86 for detecting high-risk areas.

The risk zone maps generated through GIS visualization effectively identified priority intervention areas based on environmental factors and population density. The novelty of this research lies in the integration of local epidemiological data, environmental variables, ML methods, and spatial analysis to develop a data-driven early warning system for DF outbreaks.

The results also confirmed that rainfall, humidity, air temperature, and population density significantly influence DF transmission patterns, with peak outbreaks occurring during periods of high rainfall and optimal temperatures for Aedes aegypti mosquito development.

This data-driven approach offers a practical and applicable solution to enhance community preparedness and support more targeted DF prevention strategies. For future development, it is recommended to expand the geographical coverage of the model, incorporate real-time data, and develop a web-based monitoring system to support sustainable DF prevention efforts.

Author Contributions: Conceptualization: Rosa Ratri Kusuma Hariningsih; Methodology: Rosa Ratri Kusuma Hariningsih; Software: Rosa Ratri Kusuma Hariningsih, Diwahana Mutiara Candrasari, Endang Setyawati, Syamsu Wahidin and Jevon Nataniel Putra; Validation: Rosa Ratri Kusuma Hariningsih, Diwahana Mutiara Candrasari, and Endang Setyawati;

Formal analysis: Rosa Ratri Kusuma Hariningsih, Diwahana Mutiara Candrasari, Endang Setyawati, Syamsu Wahidin and Jevon Nataniel Putra; Investigation: Rosa Ratri Kusuma Hariningsih, Diwahana Mutiara Candrasari, Endang Setyawati, Syamsu Wahidin and Jevon Nataniel Putra; Resources: Rosa Ratri Kusuma Hariningsih; Data curation: Rosa Ratri Kusuma Hariningsih; Writing—original draft preparation: Rosa Ratri Kusuma Hariningsih; Writing—review and editing: Rosa Ratri Kusuma Hariningsih, Diwahana, Syamsu Wahidin and Jevon Nataniel Putra; Visualization: Rosa Ratri Kusuma Hariningsih; Supervision: Rosa Ratri Kusuma Hariningsih; Project administration: Rosa Ratri Kusuma Hariningsih.

Funding: This research was funded by STIKOM Yos Sudarso Purwokerto.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Acknowledgments: The authors would like to express their gratitude to STIKOM Yos Sudarso Purwokerto for the financial and institutional support provided during this research..

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] S. A. Kularatne and C. Dalugama, "Dengue infection: Global importance, immunopathology and management," Clinical Medicine, Journal of the Royal College of Physicians of London, vol. 22, no. 1, pp. 9–13, Jan. 2022, doi: 10.7861/clinmed.2021-0791.

[2] F. P. Rocha and M. Giesbrecht, "Machine learning algorithms for dengue risk assessment: a case study for São Luís do Maranhão," Computational and Applied Mathematics, vol. 41, no. 8, Dec. 2022, doi: 10.1007/s40314-022-02101-z.

[3] R. Indawati, L. Y. Hendrati, and S. Widati, "The Early Vigilance of Dengue Hemorrhagic Fever Outbreak in the Community," Jurnal Kesehatan Masyarakat, vol. 16, no. 3, pp. 366–376, Mar. 2021, doi: 10.15294/kemas.v16i3.24114.

[4] Z. A. Hadi and N. C. Dom, "Development of machine learning modelling and dengue risk mapping: A concept framework," in IOP Conference Series: Earth and Environmental Science, Institute of Physics, 2023. doi: 10.1088/1755-1315/1217/1/012038.

[5] I. A. P. Salsabila, A. Santjaka, and N. Utomo, "DHF Endemicity and Aedes aegypti Larvae Density Mapping in West Purwokerto Community Health Center's Working Area in 2023," Jurnal Kesehatan Lingkungan Indonesia, vol. 23, no. 2, pp. 137–145, 2024, doi: 10.14710/jkli.23.2.137-145.

[6] O. R. Pinontoan, O. J. Sumampouw, J. H. V. Ticoalu, J. E. Nelwan, E. C. Musa, and J. Sekeeon, "The variability of temperature, rainfall, humidity and prevalance of dengue fever in Manado City," Bali Medical Journal, vol. 11, no. 1, pp. 81–86, 2022, doi: 10.15562/bmj.v11i1.2722.

[7] M. Bari Antor et al., "A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease," J Healthc Eng, vol. 2021, 2021, doi: 10.1155/2021/9917919.

[8] M. S. Rahman et al., "Mapping the spatial distribution of the dengue vector Aedes aegypti and predicting its abundance in northeastern Thailand using machine-learning approach," One Health, vol. 13, Dec. 2021, doi: 10.1016/j.onehlt.2021.100358.

[9] S. A. Arhin and A. Gatiba, "Predicting crash injury severity at unsignalized intersections using support vector machines and naïve Bayes classifiers," Transportation Safety and Environment, vol. 2, no. 2, pp. 120–132, Jun. 2020, doi: 10.1093/tse/tdaa012.

[10] N. Shahzad, X. Ding, and S. Abbas, "A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan," Applied Sciences (Switzerland), vol. 12, no. 5, Mar. 2022, doi: 10.3390/app12052280.

[11] Z. Su, L. Xu, J. Xu, J. Li, and M. Huangfu, "SIG: Speaker Identification in Literature via Prompt-Based Generation," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 17, pp. 19035–19043, 2024, doi: 10.1609/aaai.v38i17.29870.

[12] S. Bahri, "Pengembangan Model Prediksi Banjir Menggunakan Data Hidrologi dan Sistem Informasi Geografis (SIG)," WriteBox, vol. 1, no. 2, pp. 1–11, 2024, [Online]. Available: https://writebox.cloud/index.php/wb/article/view/53

[13] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[14] G. Gupta et al., "DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms," Diagnostics, vol. 13, no. 6, Mar. 2023, doi: 10.3390/diagnostics13061093.

[15] S. Fadli, M. Ashari, P. Studi Sistem Informasi, and S. Lombok, "JISA (Jurnal Informatika dan Sains) Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results," 2021.

[16] A. K. Abed, "Utilizing Artificial Intelligence in Cybersecurity : A Study of Neural Networks and Support Vector Machines," vol. 2025, pp. 14–24, 2025.

[17] A. Arista, "Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19," Sinkron, vol. 7, no. 1, pp. 59–65, Jan. 2022, doi: 10.33395/sinkron.v7i1.11243.

[18] K. N. Khikmah, I. Indahwati, A. Fitrianto, E. Erfiani, and R. Amelia, "Backwards Stepwise Binary Logistic Regression for Determination Population Growth Rate Factor in Java Island," Jambura Journal of Mathematics, vol. 4, no. 2, pp. 177–187, Jun. 2022, doi: 10.34312/jjom.v4i2.13529.

[19] S. Bhatia and J. Malhotra, "Naïve bayes classifier for predicting the novel coronavirus," in Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021, Institute of Electrical and Electronics Engineers Inc., Feb. 2021, pp. 880–883. doi: 10.1109/ICICV50876.2021.9388410.

[20] S. I. Fallo, "ABSTRACT SUPPORT VECTOR MACHINE, NA¨IVENA¨ NA¨IVE BAYES CLASSIFIER AND ORDINAL LOGISTIC REGRESSION IN WEATHER PREDICTION." [Online]. Available: http://etd.repository.ugm.ac.id/

[21] P. Bach, M. S. Kurz, V. Chernozhukov, M. Spindler, and S. Klaassen, "DoubleML: An Object-Oriented Implementation of Double Machine Learning in R," J Stat Softw, vol. 108, no. 3, pp. 1–56, 2024, doi: 10.18637/jss.v108.i03.

[22] R. Anagora, R. Taufiq, A. Dedi Jubaedi, R. Wirawan, and A. Syah Putra, "The Classification of Phishing Websites using Naive Bayes Classifier Algorithm," International Journal Of Science, [Online]. Available: http://ijstm.inarah.co.id

[23] T. Ige, C. Kiekintveld, A. Piplai, A. Waggler, O. Kolade, and B. H. Matti, "An investigation into the performances of the Current state-of-the-art Naive Bayes, Non-Bayesian and Deep Learning Based Classifier for Phishing Detection: A Survey," 2024, [Online]. Available: http://arxiv.org/abs/2411.16751

[24] A. T. Owolabi, K. Ayinde, J. I. Idowu, O. J. Oladapo, and A. F. Lukman, "A New Two-Parameter Estimator in the Linear Regression Model with Correlated Regressors," J Stat Appl Probab, vol. 11, no. 2, pp. 499–512, May 2022, doi: 10.18576/jsap/110211.

[25] M. W. A. Ashraf, A. R. Singh, A. Pandian, R. S. Rathore, M. Bajaj, and I. Zaitsev, "A hybrid approach using support vector machine rule-based system: detecting cyber threats in internet of things," Sci Rep, vol. 14, no. 1, pp. 1–19, 2024, doi: 10.1038/s41598-024-78976-1.