IJISCS (International Journal of Information System and Computer Science)
IJISCS, Volume 10, Number 1, April 2026
E ISSN: 2598-246X; P ISSN: 2598-0793, pp. 1-8
Accredited SINTA 4 Number 79/E/KPT/2023
https://jurnal.ftikomibn.ac.id/index.php/ijiscs

IMPLEMENTATION OF NAIVE BAYES ALGORITHM FOR PREDICTION OF THE SPREAD OF DENGUE HEMORRHAGIC FEVER

*Corresponding author
irmarosita26@gmail.com

Irma Rosita, Muhamad Muslihudin
Prodi Sistem Informasi, Institut Bakti Nusantara, Lampung
Jl. Wisma Rini, No.09 Pringsewu, Lampung
Email : irmarosita26@gmail.com, mmuslihudin415@gmail.com

Article history:
Received : Feb 22, 2026
Revised : March 16, 2026
Accepted : March 23, 2026

Keywords:
Naive Bayes, Dengue Hemorrhagic Fever, Data Mining, Disease Prediction, Classification

Abstract
Dengue Hemorrhagic Fever (DHF) remains a significant public health problem in Indonesia, particularly in regions with environmental conditions that support the growth of *Aedes aegypti* mosquitoes. This study aims to implement the Naive Bayes algorithm to predict the distribution of DHF cases in the working area of Wates Public Health Center, Pringsewu Regency, Lampung. The research employs a quantitative approach using data mining techniques, including data collection, preprocessing, modeling, and evaluation stages. The dataset consists of 260 DHF cases in 204-2025 categorized into acute, subacute, and non-acute conditions. The results indicate that the Naive Bayes algorithm achieved an accuracy of 65.8%, with a recall value of 0.658 and an F1-score of 0.556. The analysis also shows that most cases fall into the acute category, and there is no significant relationship between gender and disease severity. Although the AUC and MCC values are relatively low, the model provides an initial insight into the distribution pattern of DHF. Therefore, the Naive Bayes algorithm can be utilized as a decision support system to determine priority areas for intervention and as an early warning system to improve the effectiveness of DHF control.

1.0. INTRODUCTION

Dengue Hemorrhagic Fever (DHF) is an acute infectious disease caused by *the dengue virus* and transmitted to humans through the bite of *the Aedes aegypti mosquito*. According to data from the Ministry of Health of the Republic of Indonesia (Kemenkes RI, 2024), Dengue Hemorrhagic Fever (DHF) in Indonesia in 2021-2025 has seen a significant increase and decrease in cases[1]. In 2021 there were 73,518 cases. The number of these cases increased in 2022 to 143,266 cases. In 2023, it decreased compared to the previous year to 114,720 cases. This happened because of the intensive control of the spread of Dengue Hemorrhagic Fever (DHF) towards the end of the year. However, in 2024 Dengue Hemorrhagic Fever (DHF) cases will increase sharply with a total of 119,709 cases, this number of cases is higher than the previous year's period, namely 2023. The increase in cases began at the end of February 2024,

driven by weather factors. It was recorded that 124 patients died until the 8th week of 2024. In 2025, Dengue Hemorrhagic Fever (DHF) cases will increase rapidly at the beginning of the year. The Ministry of Health reported a total of 131,393 cases of Dengue Hemorrhagic Fever (DHF) during January-October 2025 with 544 deaths. The factors causing the 2024-2025 surge are caused by climate change affecting the life cycle of *Aedes aegypti* mosquitoes, with a combination of weather such as erratic rainfall and drought that causes mosquito breeding, as well as a lack of environmental cleanliness that triggers the growth of mosquito larvae. Based on data from the Ministry of Health of the Republic of Indonesia, West Java Province is the province with the highest cases of Dengue Hemorrhagic Fever (DHF). Data from 2024-2025 shows an increase in cases of almost 3 times compared to the previous year. In this case, the age vulnerability of being infected by Dengue Hemorrhagic Fever (DHF) is in the age group of 15-44 years, but the mortality rate caused by this disease ranges from children (5-14 years). The cause of Dengue Hemorrhagic Fever (DHF) is climate change (*global warming*) the main symptom that is often shown Dengue Hemorrhagic Fever (DHF) is a high fever close to 40 °C for 2-7 days, then on the 4-5th day there is a decrease in body temperature then there is an increase, body temperature accompanied by joint pain, headache, pain behind the eyes, red spots, and nausea[1][2].

This also happens in the working area of the Wates Health Center of Pringsewu Regency, where every year there are different cases of Dengue Hemorrhagic Fever (DHF). It was recorded that in 2021-2025 there were 1,000 cases caused by Dengue Hemorrhagic Fever (DHF). This disease tends to appear often every year with a diverse distribution pattern that is almost difficult to predict, with the changing weather around the Wates Health Center's work area making the surrounding community infected by Dengue Hemorrhagic Fever (DHF). Dengue Hemorrhagic Fever (DHF) is still one of the problems being faced, where Pringsewu Regency, is one of the areas of Lampung Province that has experienced a significant increase in Dengue Hemorrhagic Fever (DHF) cases.

Based on data obtained from the Pringsewu Regency Health Office (Dinkes, 2023), cases of Dengue Hemorrhagic Fever (DHF) have increased significantly in recent years. In 2021 the number of cases was recorded as 192, then increased to 279 cases in 2022 with three deaths. The sharpest increase will occur in 2024 with a total of 723 cases. The highest increase in cases occurred in February and March 2024, while the lowest number of cases was recorded in November. The surge in cases is influenced by several factors, including the increasing population of *Aedes aegypti* mosquitoes during the rainy season and the low level of public awareness in maintaining environmental cleanliness and carrying out mosquito nest eradication activities (PSN). Dengue Hemorrhagic Fever (DHF) is an infectious disease that is still a serious problem in Indonesia, with the rate of cases fluctuating every year due to environmental factors and climate change. The significant increase in cases, including in Pringsewu Regency, shows that the spread of this disease is difficult to predict conventionally. This condition also occurs in the work area of the Wates Health Center, where dengue cases continue to appear every year with an erratic pattern.

The countermeasures that have been carried out, such as fogging and health education, have not been fully effective in reducing the number of cases. This is due to the limitations of conventional methods that have not been able to analyze the distribution pattern thoroughly. Therefore, an information technology-based approach, especially data mining, is needed to help predict the spread of diseases more accurately and systematically. Data mining is a technique that is able to extract hidden patterns from large amounts of data, making it very relevant to be used in the health field. One of the widely used methods is the Naive Bayes algorithm, which has advantages in probability-based classification and is able to produce fast and accurate predictions. This method can be used to classify the risk level of dengue spread based on time, location, and environmental conditions. Previous research has shown that data mining techniques, both using clustering and decision tree methods, have proven to be effective in predicting and classifying dengue cases with a high level of accuracy. However, the application of the Naive Bayes algorithm for predicting the distribution of dengue at the health center level, especially in the Pringsewu area, is still limited. This is a research opportunity to present a new approach that is more specific and contextual. Based on these problems, this study aims to implement the Naive Bayes algorithm in predicting the spread of dengue disease in the working area of the Wates Health Center. The resulting model is expected to be able to

provide accurate prediction information as a basis for decision-making, determination of priority areas, and as an early warning system in efforts to prevent and control diseases more effectively, efficiently, and data-based.

2.0. RESEARCH METHODS

2.1. Data Collection Techniques

The data collection technique in this study was carried out through the observation method, namely by directly observing Dengue Hemorrhagic Fever (DHF) case data available at the Wates Health Center. The observation was focused on data for 2024-2025 which recorded as many as 260 cases of dengue with variations in patient conditions. The data was analyzed based on disease status, namely acute, sub-acute, and non-acute dengue. Acute status indicates the condition of the patient with severe symptoms and requires intensive treatment, sub-acute describes a transitional condition with moderate severity, while non-acute indicates a mild condition or early stage of the disease. This observation aims to understand the pattern of case distribution as well as the characteristics of each disease category. In addition to observation, interview techniques were carried out with health workers at the Wates Health Center, such as doctors and surveillance officers, to obtain more in-depth information related to the causative factors and patterns of dengue spread. This interview explores information about the relationship between environmental conditions, weather changes, community behavior, and population density to the increase in dengue cases. In addition, interviews are also used to understand the classification of disease status (acute, sub-acute, and non-acute) based on the patient's clinical condition, so that the data obtained is not only numerical but also has a strong medical context. The next technique is documentation, namely the collection of historical data from health reports, patient medical records, and archives of the Pringsewu Regency Health Office. This documentation data is used to complement the results of observations and interviews, as well as as a basis for the data mining analysis process. By combining these three techniques, this study ensures that the data used has a high level of validity, accuracy, and relevance. Case data for 2024-2025 with 260 dengue patients who have been classified based on severity is the main foundation in building a prediction model using the Naive Bayes algorithm.

2.2. Algoritma Naïve Baiyes

[3] [4][2], [5], [6] Naïve Bayes is a technique for predicting future possibilities based on previous experience by referring to probabilities and statistics. [7], [8] Naive Bayes Method is one of the popular classification methods in the field of data mining and machine learning. This method is based on Bayes' probability theorem with simple and naïve assumptions that make it easy to implement and efficient in data analysis. [9] [10] *Naive Bayes Classifier* (NBC) is a supervised classification method that is based on Bayes' theorem derived from conditional probability, the Naive Bayes Classifier is widely used and has evolved into five advanced categories.

Based on the opinions of these experts, it can be concluded that Naïve Bayes is a supervised classification method used to predict or predict an event based on previous data or experience by utilizing the concepts of probability and statistics. This method relies on Bayes' theorem assuming independence between attributes, so that each variable is considered to be incompatible with each other. This simple assumption makes the calculation process easier and more efficient in data processing. Therefore, Naïve Bayes is widely applied in the field of data mining and machine learning and continues to be developed into various forms further according to the needs of analysis. Naïve Bayes is one of the classification algorithms in the *supervised learning* approach that is widely used in data mining and *machine learning* because it is based on Bayes' Theorem in probability theory to estimate the probability of an event based on existing evidence. In the classification process, this algorithm determines the chances of a data being included in a certain class by calculating the posterior probability using the formula:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad \text{..... (1)}$$

Description:

1. P(H | X) = posterior probability, i.e. the probability of a class H after considerin the data or proof X,
2. P(X | H) = *likelihood*, i.e. the probability of the occurrence of data X if class H is known,
3. P(H) = *prior probability*, i.e. the initial probability of a class before paying attention to the data,
4. P(X) = *evidence*, which is the overall probability of the data X used to normalize the calculation results[11], [12], [13].

3.0. DISCUSSION

3.1. Data Mining Testing

3.1.1. Workflow view

The *workflow* display in the image shows the data analysis process using the data mining method with the Naive Bayes algorithm to predict the spread of Dengue Hemorrhagic Fever (DHF). The process starts from data input through the *File* component, then the *Preprocess* stage is carried out to clean and prepare the data so that it is ready for analysis. Furthermore, in the *Select Columns* stage, important attributes were selected according to the study variables, such as time, location, and patient condition (acute, sub-acute, and non-acute). The processed data is then entered into the Naive Bayes model for a probability-based classification process. After the modeling process, an evaluation is carried out using the *Test and Score feature* to measure the performance of the model, the results of which are visualized through *the Confusion Matrix* to determine the level of accuracy, precision, and recall. In addition, the prediction results are also displayed in the form of *Predictions*, *Pivot Tables*, and *Box Plots* to provide an overview of the data distribution and classification results. This workflow shows that the research has used a systematic approach based on data mining to produce an accurate model for predicting the distribution of dengue, so that it can support decision-making at the Wates Health Center more effectively and data-based.

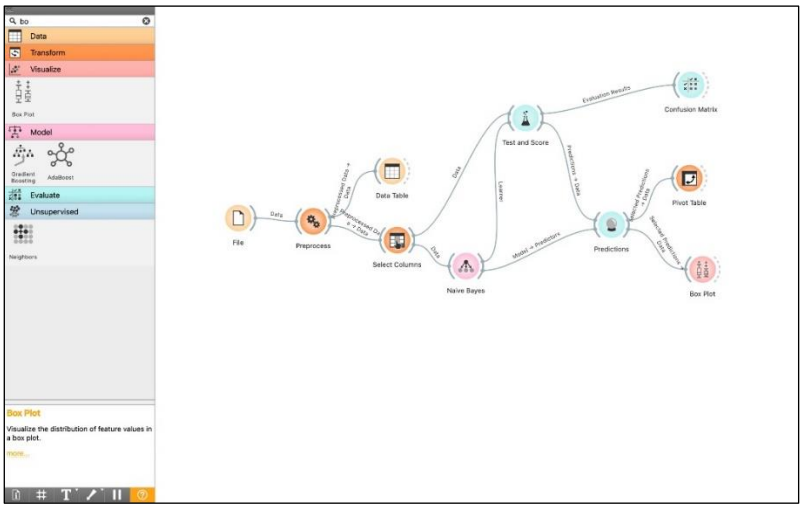

Figure 1. Workflow view

1. ***Pivot Table* from *the Naive Bayes model***

   The results of the *Pivot Table analysis from the* Naive Bayes *model* in predicting the spread of Dengue Hemorrhagic Fever (DHF) disease based on gender variables (JK) and disease status (acute, subacute, and non-acute). It can be seen that the number of cases in men (L) is 118 cases with the dominance of the acute category (83 cases), while women (P) have a higher number of cases of 142 cases with 97 cases included in the acute category. Overall, the total data showed 260 cases with the largest distribution in the acute (180 cases) category, followed by non-acute (50 cases) and subacute (30 cases). The mean probability value of Naive Bayes showed the highest prediction tendency in the subacute category (0.648), which indicates that the model was able to identify classification patterns based on the distribution of patient data. These results support the research

objectives in implementing the Naive Bayes algorithm to predict the distribution of dengue in the Wates Health Center more accurately, as well as provide important information in determining treatment priorities based on patient characteristics.

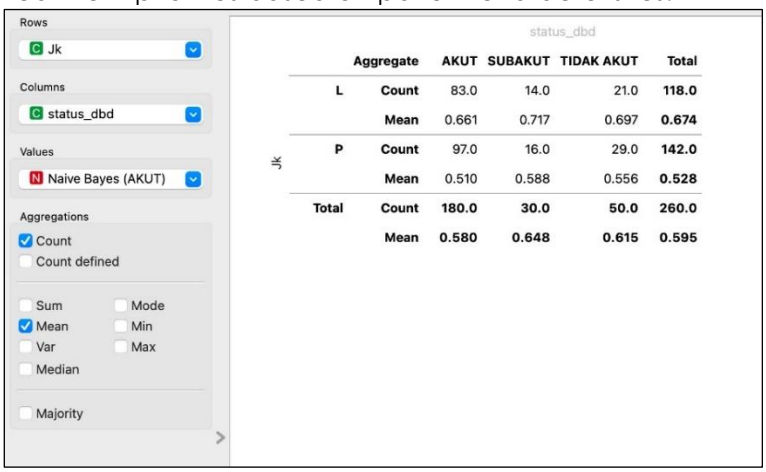

Figure 2. *Pivot Table* from *the Naive Bayes model*

2. Visualization of the distribution of the results *of the Naive Bayes Algorithm classification*

Visualization of the distribution of the results of *the Naive Bayes algorithm* classification on Dengue Hemorrhagic Fever (DHF) disease status by sex (L and P). The bar chart shows that in both males and females, the acute dengue category predominates over subacute and non-acute, which is evident from the largest blue bar length. Meanwhile, the subacute (red) and non-acute (green) categories have a smaller proportion, but still show variations in distribution in each group. The results of the Chi-Square statistical test ($\chi^2 = 0.29$ with p-value = 0.865) showed that there was no significant relationship between sex and dengue status, so it can be concluded that the prevalence of dengue severity is relatively even between men and women. These findings strengthen the results of the study that the Naive Bayes model is able to objectively identify the distribution pattern of dengue cases and can be used as a basis for predicting disease spread prediction.

Figure 3. Visualization of the distribution of the results *of the Naive Bayes algorithm classification*

3. Model Evaluation Using Test *and Score Feature*

The results of the model evaluation using *the Test and Score* feature in the Orange application show the performance of *the Naive Bayes* algorithm in predicting the spread of Dengue Hemorrhagic Fever (DHF). The test was carried out with *a cross validation* method of 10 folds, so that the model was tested repeatedly to obtain more stable and objective results. Based on the results of the evaluation, an *Accuracy value (CA) of 0.658*

was obtained, which shows that the model is able to classify data with an accuracy level of around 65.8%. This value indicates that the Naive Bayes model is quite good at predicting the categories of dengue status (acute, subacute, and non-acute). A *Precision value of 0.491*, *a Recall of 0.658*, and *an F1-Score of 0.556* indicate the balance of model performance in identifying positive data and minimizing misclassification. The high recall value indicates that the model has a good ability to detect actual dengue cases, so it can support early warning systems. However, the precision value that is still below 0.5 indicates that there are still some errors in classification, so the model needs further development, for example through the addition of attributes or data optimization. The AUC value of 0.375 and the MCC of -0.057 indicate that the model's ability to distinguish classes is still relatively low. This can be due to a limited amount of data, class imbalances, or a lack of relevant variables in the dataset. Nevertheless, these results still show that the Naive Bayes algorithm is able to provide an initial picture of the distribution pattern of dengue in the Wates Health Center. Therefore, the model can still be used as a basis for decision-making, with further development recommendations to improve prediction accuracy and performance.

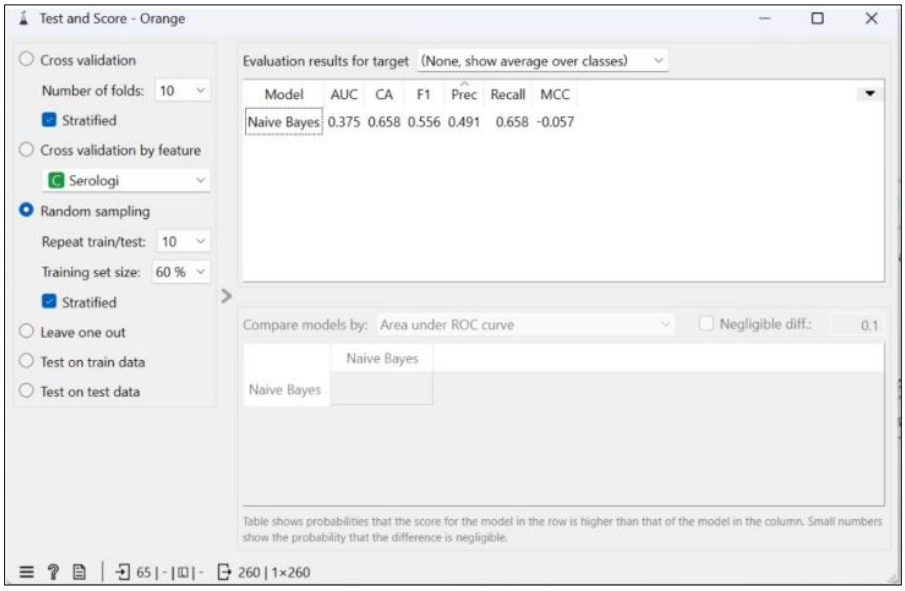

Figure 4. Model evaluation using *the Test and Score feature*

3.2. Discussion

The results of the study showed that *the Naive Bayes algorithm* was able to classify the distribution of Dengue Hemorrhagic Fever (DHF) in the working area of the Wates Health Center with an accuracy rate of 65.8%. This value indicates that the model has a fairly good ability to predict the categories of dengue status (acute, subacute, and non-acute) based on the variables used. The distribution of data shows the dominance of cases in the acute category, which is in line with the epidemiological conditions of dengue which tend to increase in certain seasons. This reinforces that the data mining approach can provide a more systematic picture of the pattern of disease spread than conventional methods. Based on the results of the model evaluation, a *recall* value of 0.658 shows that the model is quite effective in detecting dengue cases that actually occur, so it has the potential to be used as an early warning system. However, a *precision* value of 0.491 indicates that there are still misclassifications that need to be corrected. This condition is generally caused by the limitations of the attributes of the data used, such as the lack of environmental variables and social factors. These findings are in line with research by [14] which stated that the performance of the Naive Bayes model in disease prediction is greatly influenced by the quality and completeness of the data used, especially in the case of environment-based infectious diseases.

The results of the data distribution visualization showed that there was no significant difference between sex and the severity of dengue, which was proven by the Chi-Square test

with a significance value (p-value > 0.05). This shows that gender factors are not the main determinants in the spread of dengue, but are more influenced by environmental factors such as rainfall, temperature, and population density. These findings are consistent with the research of [5] which states that the spread of dengue is more influenced by ecological and climatic factors than individual characteristics. Although the AUC value of 0.375 and *the MCC of -0.057* indicate that the model's discriminating ability is still low, the Naive Bayes model is still able to provide a useful initial picture of classification patterns. This limitation can be caused by a relatively limited amount of data as well as an imbalance in class distribution. Research by [15], [16] also mentioned that Naive Bayes has optimal performance on balanced datasets and has relevant attributes. Therefore, improving data quality and feature selection is an important factor in improving model performance.

This study proves that the implementation of the Naive Bayes algorithm can be used as an initial approach in predicting the distribution of dengue at the health center level. The resulting model can be used as a basis for decision-making in determining priority areas for treatment as well as as a support system in disease prevention. These findings are in line with the research [17], [18], [19] which affirm that the use of data mining and machine learning technology is very effective in supporting data-based dengue disease control. Thus, this research makes a real contribution to the development of a more adaptive, efficient, and applicative disease prediction system in the healthcare environment.

4.0. CONCLUSION

This study successfully implemented the *Naive Bayes* algorithm in predicting the spread of Dengue Hemorrhagic Fever (DHF) in the working area of the Wates Health Center, Pringsewu, Lampung. The test results showed that the model had an accuracy rate of 65.8% with a fairly good recall value, so that it was able to identify dengue cases quite effectively. The distribution of data showed that the category of acute dengue dominated compared to subacute and non-acute, and there was no significant relationship between sex and disease severity, so that environmental factors played a greater role in the spread of dengue. Although the AUC and MCC values are still relatively low, the Naive Bayes model is still able to provide an initial picture of the pattern of classification of disease distribution. This shows that this method is feasible to use as an initial approach in data-driven prediction systems. However, to improve model performance, further development is needed through the addition of more relevant variables, improved data quality, and the use of optimization methods or algorithm combinations. This research contributes to the use of data mining technology in the health sector, especially in supporting prediction systems and decision-making related to the handling of dengue. The resulting model is expected to be used as an early warning system and assist the Puskesmas in determining priority areas for handling more effectively, efficiently, and data-based.

REFERENCES

[1] Kementerian Kesehatan Republik Indonesia, *Profil Kesehatan Indonesia Tahun 2024*. Kemenkes RI, 2024.

[2] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. Ipnuwati, *Implementasi Data Mining (Clastering, Association, Estimation, Classification)*. Penerbit Adab, 2023.

[3] P. N. Tan, M. Steinbach, V. Kumar, and A. Karpatne, *Data Mining*, vol. 11, no. 1. 2019.

[4] S. Mukodimah, M. Muslihudin, D. R. Mustofa, and D. Susianto, "Naive Bayes Classifier Method Analysis and Support Vector Machine (SVM) Student Graduation Prediction," *NEUROQUANTOLOGY*, vol. 20, no. 12, pp. 3522–3533, 2022, doi: 10.14704/NQ.2022.20.12.NQ77360.

[5] S. Bhatt, P. W. Gething, O. J. Brady, and others, "The global distribution and burden of dengue," *Nature*, vol. 496, no. 7446, pp. 504–507, 2018, doi: 10.1038/nature12060.

[6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. Morgan Kaufmann, 2021.

[7] S. Mukodimah and M. Muslihudin, "The Naïve Bayes Method as a Measurement Model Effectiveness of Online Learning," *J. TAM*, vol. 13, no. 2, pp. 131–137, 2022.

[8] L. J. Anreaja, N. N. Harefa, J. G. P. Negara, V. N. H. Pribyantara, and A. B. Prasetyo, "Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea

Mobile Application Users in Indonesia," *JISA(Jurnal Inform. dan Sains)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.31326/jisa.v5i1.1267.

[9] S. Mukodimah and M. Muslihudin, "The Naïve Bayes Method as A Measurement Model Effectiveness of Online Learning," *Technol. Accept. Model. J. TAM*, vol. 13, no. 2, pp. 131–137, 2022.

[10] V. Fitriyana, L. Hakim, D. C. R. Novitasari, A. Hanif, and Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *J. Buana Inform.*, vol. 14, no. 1, pp. 40–49, Dec. 2023.

[11] Y. L. Hii, H. Zhu, N. Ng, L. C. Ng, and J. Rocklöv, "Forecast of dengue incidence using temperature and rainfall," *PLoS Negl. Trop. Dis.*, vol. 6, no. 11, p. e1908, 2020, doi: 10.1371/journal.pntd.0001908.

[12] A. H. Mahfudzin, S. Sriyanto, S. Sutedi, and W. Wasilah, "Prediction of DHF Disease Using Bagging Algorithm with Decision Tree C4.5," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 7, no. 3, pp. 747–758, 2025.

[13] A. Safitri, D. Rahmawati, and R. Putra, "Penerapan algoritma C4.5 untuk klasifikasi dini pasien Demam Berdarah Dengue (DBD)," *Nama J.*, vol. 10, no. 2, pp. 123–130, 2022.

[14] T. Chen, C. Guestrin, and T. He, "Machine learning approaches for disease prediction and classification in healthcare," *IEEE Access*, vol. 9, pp. 123456–123467, 2021.

[15] C. Liao, Y. Huang, Z. Zheng, and Y. Xu, "Investigating the factors influencing urban residents' low-carbon travel intention: A comprehensive analysis based on the TPB model," *Transp. Res. Interdiscip. Perspect.*, vol. 22, no. February, p. 100948, 2023, doi: 10.1016/j.trip.2023.100948.

[16] A. Wusiman *et al.*, "Preparation and sulfate modified of Lagenaria siceraria (Molina) Standl polysaccharide and its immune-enhancing adjuvant activity," *Poult. Sci.*, vol. 101, no. 11, pp. 1–10, 2022, doi: 10.1016/j.psj.2022.102112.

[17] A. Seah, J. Aik, L. C. Ng, and C. C. Tam, "The effects of maximum ambient temperature and heatwaves on dengue infections in the tropical city-state of Singapore – A time series analysis," *Sci. Total Environ.*, vol. 775, p. 145117, 2021, doi: 10.1016/j.scitotenv.2021.145117.

[18] O. J. Brady and S. I. Hay, "The global expansion of dengue: How aedes aegypti mosquitoes enabled the first pandemic arbovirus," *Annu. Rev. Entomol.*, vol. 65, pp. 191–208, 2020, doi: 10.1146/annurev-ento-011019-024918.

[19] C. Jing *et al.*, "Emerging Risk to Dengue in Asian Metropolitan Areas Under Global Warming," *Earth's Futur.*, vol. 12, no. 7, pp. 1–13, 2024, doi: 10.1029/2024EF004548.