Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13023
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Performance of CART Time-Based Feature Expansion in Dengue Classification Index Rate

Annisya Hayati Suhendar[1)*], Aniq Atiqi Rohmawati [2)], Sri Suryani Prasetyowati [3)]
[1,2,3)]School of Computing, Telkom Univesity, Indonesia
[1)]annisyahs@student.telkomuniversity.ac.id, [2)]aniqatiqi@telkomuniversity.ac.id,
[3)]srisuryani@telkomuniversity.ac.id

Submitted : Sep 18, 2023 | **Accepted** : Oct 15, 2023 | **Published** : Jan 1, 2024

Abstract: This study proposes utilizing the machine learning technique CART to classify the spread of dengue hemorrhagic fever (DHF). To expand the features used, the CART classification model was developed based on data collected over the previous 2 to 4 years. The data sources included the Bandung City Health Office for the cases of DHF, the Bandung Meteorology, Climatology and Geophysics Agency for the climate data, the Bandung City Central Statistics Agency for population and educational history data. The top-performing CART classification model over the past 2, 3, and 4 years achieved accuracies of 93%, 93%, and 90%, respectively. The models that exhibited the highest accuracy values and optimal number of feature extensions were chosen as the best ones. CART is among several machine learning techniques that can effectively measure the most impactful features during the classification process. The meteorological parameters were found to be irrelevant in the classification process. This study reveals that the population size, male population proportion, and educational attainment levels are the most impactful features in the classification of DHF spread in Bandung City. The research provides valuable insights into the classification of DHF spread in Bandung City through feature expansion.
Keywords: CART, Classification, Dengue Hemorrhagic Fever, Feature Expansion, Machine Learning

INTRODUCTION

The high number of cases of Dengue Hemorrhagic Fever (DHF) is still a health problem in Indonesia. At the end of 2020, DHF cases in Indonesia reached 95,893 cases with a total death of 661 people. The city of Bandung is one of the cities in Indonesia with the highest cases of 2,363 cases (Ministry of Health, 2020).

With the increase in the spread and death rates in the city of Bandung due to DHF, it is necessary to prevent its spread. Rapid spread could have been prevented if information on the number of cases was available in the future. Information for the future can be in the form of classification predictions for the number of cases that can indicate the level of vulnerability of an area to the spread of DHF. This can help the government formulate a strategy to prevent the spread of the disease. Climatic factors, such as alterations in precipitation, temperature, humidity, and wind patterns, have a profound impact on both terrestrial and marine ecosystems, ultimately affecting the well-being of all organisms, including human beings. In addition, there exist ecological factors that contribute to the spread of dengue fever, often complicated by the lack of community vigilance in preventing outbreaks (Gui, Gwee, Koh, & Pang, 2021).

Over the past 12 years, the Incident Rate (IR) has tended to fluctuate drastically. In 2017, there were 68,407 cases of dengue hemorrhagic fever (DHF) reported with an incidence rate of 26.12 per 100,000 individuals. It appears that the incidence rate decreased in 2017 compared to 2016, when there were

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

204,171 cases and an incidence rate of 78.85 per 100,000 population (Data and Information Center of the Indonesian Ministry of Health, 2018).

There are several studies that discuss the prediction and classification of the spread of DHF conducted by previous researchers. Research (Inayah, Prasetyowati, & Sibaroni, 2020), used the Hybrid Classification method, namely Naïve Bayes, K-Nearest Neighbor and Artificial Neural Networks to produce an accuracy value of up to 90%. This shows that the hybrid method produces higher accuracy in predicting the spread of the disease.

Meanwhile, (Wahono & Riana, 2020) study aims to predict potential blood donors through the comparative analysis of multiple algorithms. The research compares the highest accuracy values among the Naïve Bayes, K-Nearest Neighbor, and Decision Tree C4.5 algorithms. Based on the performance test results of the three algorithms, the Decision Tree C4.5 algorithm demonstrates the highest accuracy rate with an accuracy value of 93.83%. Compared to the Naïve Bayes and K-Nearest Neighbor algorithms, which achieved accuracy values of 85.15% and 84.10%, respectively.

In research which implemented the SVM, Naïve Bayes and Random Forest methods, the accuracy values were 77.5%, 56%, and 84.1%, respectively. From the comparison of the three methods, Random Forest has the best classification method (Arafiyah, Hermin, Kartika, Alimuddin, & Saraswati, 2018).

Similar studies predicting the spread of dengue (Astuti et al., 2016) stated that the application of the J48 algorithm resulted in an accuracy value of 75.83% using cross validation fold 5 and an accuracy value of 80% using fold 10.

Another study used a combination of algorithms in the dengue fever classification process. This study uses the Naïve Bayes and Decision Tree hybrid method, and produces a high accuracy value, which is 92% (Taneja & Gautam, 2019).

However, there has been no research that has considered the co-factors for the spread of DHF as mentioned in (Center for Epidemiological Data and Surveillance, 2010) using Classification and Regression Tree (CART) algorithm based on time. Thus, this study proposes the method for the classification on the spread of DHF and compare the accuracy with other methods.

LITERATURE REVIEW

According to (World Health Organization, 2023), Dengue Hemorrhagic Fever is now regarded as a global burden, with a rapid spike in cases from 505,430 in 2000 to 5.2 million in 2019.

Dengue transmission is linked to several social and environmental factors, including population density, population knowledge, and climate change in tropical and subtropical regions. (World Health Organization, 2023).

A study conducted by Ye Guoguo, Zhixiang Xu, and Minghui Yang (Ye et al., 2023) revealed that the majority of patients infected with the dengue virus (DENV) are male. Furthermore, the study found that temperature shifts may enhance the spread of dengue hemorrhagic fever (DHF) and increase the risk of an epidemic.

A study presents a visualization map and forecasts for effective management of dengue cases using fuzzy techniques through Geographic Information System and Exponential Smoothing models with an accuracy of 55%. The analysis covers data from 2012 to 2017 and employs the variables of DHF patient counts, population density, and rainfall (Mufid et al., 2018).

Another study demonstrates that factors such as the number of DHF cases and deaths, population, and rainfall are critical in calculating vulnerability values (Fariza, Mu'Arifin, & Astuti, 2021). However, this study disregards other important meteorological parameters such as temperature, humidity, and wind speed, as noted in (Asadi, Trinugroho, Hidayat, Rahutomo, & Pardamean, 2022).

A comparison study was conducted to predict the spread of dengue disease using various techniques, including Naïve Bayes, REP Tree, Random Tree, J48, and Sequential Minimal Optimization (SMO). The results showed that J48 and Naïve Bayes were the top-performing classifiers, achieving 92% and 88% accuracy, respectively (Shaukat Dar & Ulya Azmeen, 2015).

Another study employed five machine learning techniques, namely Extra Tree Classifier (ETC), eXtreme Gradient Boosting (XGB), K-Nearest Neighbor (KNN), Gradient Boosting Classifier (GBC), and Light Gradient Boosting Machine (LightGBM), to predict dengue disease. The study concluded that

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13023
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the ETC classifier exhibited the greatest accuracy, achieving 99.12% (Abdualgalil, Abraham, & M. Ismael, 2022).

A study (Hamdani, Hatta, Puspitasari, Septiarini, & Henderi, 2022) utilized Support Vector Machines (SVM) and Cross-Validation Techniques to classify dengue types. The results showed that SVM, with k-fold values of 3 and 10, achieved the highest accuracy rate of 99.1%.

METHOD

This study will develop a classification system for the spread of Dengue Hemorrhagic Fever based on expanded features from the past 2 to 4 years, using the Classification and Regression Tree (CART) method. The system design is illustrated in Figure 1.

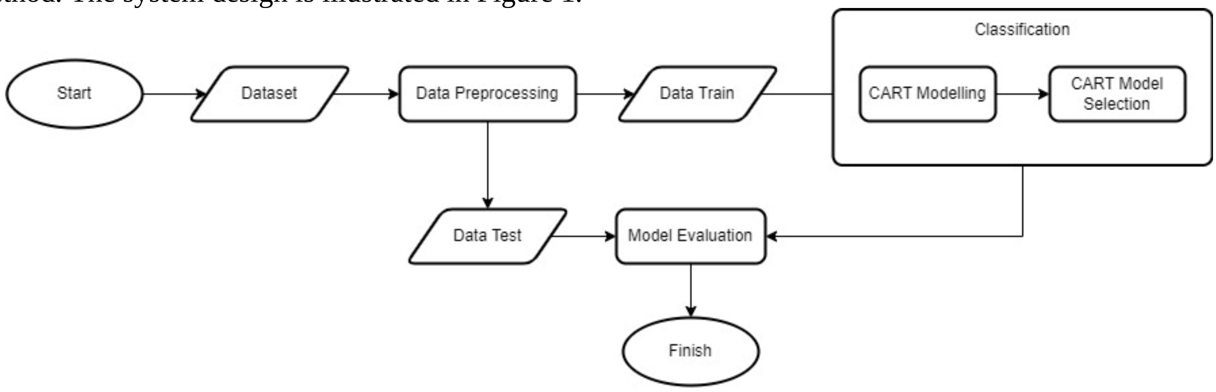

Fig. 1 System Design

Dataset

This study utilizes data on cases of dengue hemorrhagic fever (DHF) acquired from the Bandung City Health Office, population data and educational history data from the Bandung City Central Statistics Agency, and climate data from the Bandung Meteorology, Climatology, and Geophysics Agency. The data collection spanned from 2017 to 2021 and involved 30 sub-districts within Bandung City. There are 13 features represented by notation X1... X(n), with each feature notated and described as shown in Table 1.

Table 1. Dataset

Notation	Description
X1	Overall Population
X2	Proportion of Male Population
X3	Primary School Graduate
X4	Secondary School Graduate
X5	High School Diploma
X6	University Graduate
X7	Rainfall (mm)
X8	Temperature (°C)
X9	Humidity (%)
Y	Incidence rate (per 100,000 of the population)

Data Preprocessing

As the data obtained is still raw, data preprocessing is necessary to get good quality data so it can be processed to build a classification model.

The Incident Rate (IR) indicates the number of cases of DHF per 100,000 individuals in the population (Harapan, Michie, Mudatsir, Sasmono, & Imrie, 2019) with a formula:

$$IR = \frac{Case}{Population} x\ 100.000 \qquad (1)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13023
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 2 displays the case categories based on the number of incidents as per the Center for Epidemiological Data and Surveillance (2010). An area is considered low risk if the IR is below 55. Medium risk areas are identified if the IR falls between 55 and 100, whereas high-risk areas are those where the IR exceeds 100.

Table 2. Class Labeling

Class	Class Label	Range
Low	0	IR < 55
Medium	1	55 < IR < 100
High	2	IR > 100

To reduce potential biases, this study employs the stratified k-fold cross-validation technique to divide the data into two portions: training and test data (Widodo, Brawijaya, & Samudi, 2022). The dataset is divided into multiple folds based on the value of k, with each fold undergoing a training process and model testing (Zeng, Jiang, & Chen, 2019). A value of k=10 is utilized in this study due to the limited dataset size.

Classification and Regression Tree (CART) Model

The way the CART algorithm works involves classification and regression processes. The regression process involves the existing attributes, and the classification uses a decision tree. First, a recursive separation of the sample values is performed. Then, a truth tree is formed from this algorithm and will search all variables to get the most optimal value based on the goodness value (Timofeev, 2004).

In building a classification tree, there are three stages: choosing a sorter, determining the terminal node, and marking the class label. In the sorting section, the impurity value or Gini index will be used to determine the heterogeneity of a node and the selected node is the node that has the highest degree of homogeneity.

To find the value of goodness, the following equation is used:

$$\phi(s,t) = \Delta i(s,t) = i(t) - P_L i(t_L) - P_R i(t_R) \qquad (2)$$

Where $\phi(s,t)$ is the criterion of goodness of split, $i(t)$ is the Gini index heterogeneity function, $P_L i(t_L)$ is the proportion of t-node observations towards the left node, and $P_R i(t_R)$ is the proportion of t-node observations towards the right node. The equation of $i(t)$ is as follows:

$$i(t) = \sum_{i \neq j} p(i|t)p(j|t) \qquad (3)$$

$p(i|t)$ is the proportion of class i at node t, and $p(j|t)$ is the proportion of class j at node t. After the sorter is obtained, look for the terminal node. The terminal node is considered unimportant if there is no prediction class that is significant to the actual class or has reached the maximum limit of the tree. If these conditions are met, then the t node is not sorted but becomes a terminal node (Timofeev, 2004). Figure 2 shows the structure of CART.

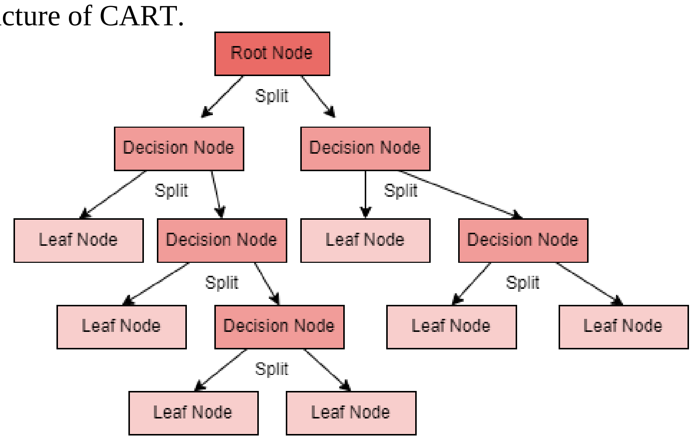

Fig. 2 Classification and Regression Tree Structure

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13023
e-ISSN : 2541-2019
p-ISSN : 2541-044X

This study developed the CART model by expanding on features from previous years. Table 3 illustrates how data collected from the past 5 years can be expanded, using the previous 2, 3, and 4 years, with the current year's target as the objective.

Table 3. Scenario of prediction model feature expansion

Model	Label	Data Training Feature	Target Prediction
2 years before	2A	2019, 2020	2021
	2B	2018, 2019	2020
	2C	2017, 2018	2019
3 years before	3A	2018, 2019, 2020	2021
	3B	2017, 2018, 2019	2020
4 years before	4A	2017, 2018, 2019, 2020	2021

To predict 2020 based on the expansion of data features from 2017, 2018, and 2019, feature expansion scenarios were considered. Table 4 displays examples of feature expansion combinations used in previous years.

Table 4. Example of feature expansion combination from the previous three years

Number of Features	Features Combination
3	X3, X12, X21
3	X3, X12, X21
4	X3, X4, X12, X21
4	X3, X4, X12, X21
5	X3, X4, X12, X13, X21
5	X3, X4, X12, X13, X21
…	…
27	X1, X2, X3, …, X25, X26, X27

CART Model Selection

The optimal CART prediction model is chosen based on its maximum accuracy value and the amount of necessary feature extensions. Accuracy is determined by calculating the total predicted accuracy for each document divided by the total number of predictions classified within the classes. (Ghoneim, 2019).

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (4)$$

TP (True Positive) occurs when the predicted category is positive, but the actual category is negative. TN (True Negative) occurs when the predicted category is negative, and the actual category is also negative. FP (False Positive) occurs when the predicted category is positive, but the actual category is negative. FN (False Negative) occurs when the predicted category is negative, but the actual category is positive. (Taneja & Gautam, 2019).

Meanwhile, the optimal number of feature extensions is determined with Sklearn SelectKBest. This method conducts a univariate statistical analysis on each variable, selecting only the few k features with the highest scores. (Irmanita et al., 2021).

RESULT

This study used the classification and regression tree method to build the classification model with the expansion of features from 2 to 4 years before. The accuracy score and feature expansion are key factors in selecting the optimal CART model, and the highest accuracy value plays a critical role in evaluating the model. Table 5 displays the accuracy value for each developed model.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13023
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 5. Prediction Model Results

Model	Label	Accuracy
2 years before	2A	87.29%
	2B	90%
	2C	89.38%
	Combined	88.89%
3 years before	3A	87.87%
	3B	89.07%
	Combined	88.47%
4 years before	4A	87.94%

Table 5 illustrates that Model 2B, with a value of 90%, outperforms other 2-year models, including the combined Model of 2A, 2B, and 2C. As for the 3-year models, model 3B has the highest accuracy of 89.07%. Meanwhile, the 4-year model only has one model with the value of the accuracy 87.94%. None of the combined models is the superior prediction model, although the combined 2-year model outperforms model 2A, and the combined 3-year model outperforms model 3A.

To calculate the accuracy of the developed model, feature expansion is performed. The result of the 2-year model accuracy is shown in Figure 2. In this model, we test feature expansion by expanding the range of features to 3-18 features. Model 2B with the expansion of 7 and 9 features, as well as model 2C with the expansion of 14 features has the highest accuracy value of 93.33%.

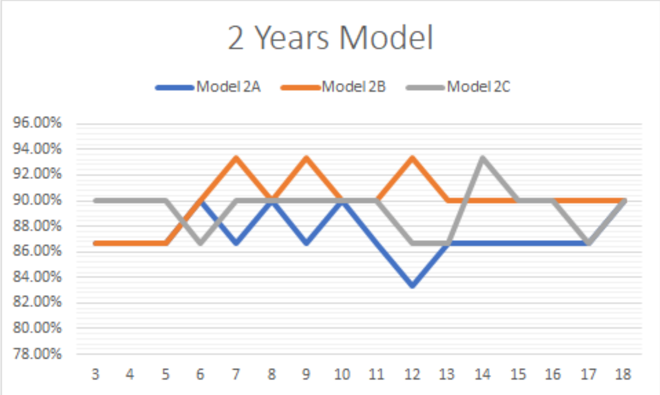

Fig. 3 Accuracy of the 2-year models

The result of the 3-year model accuracy is shown in Figure 3. In this model, feature expansion is performed by testing a range of 3-27 features for expansion. Model 3A with the expansion of 9 features and model 3B with the expansion of 11 features are the best expansion, with an accuracy of 93.33%.

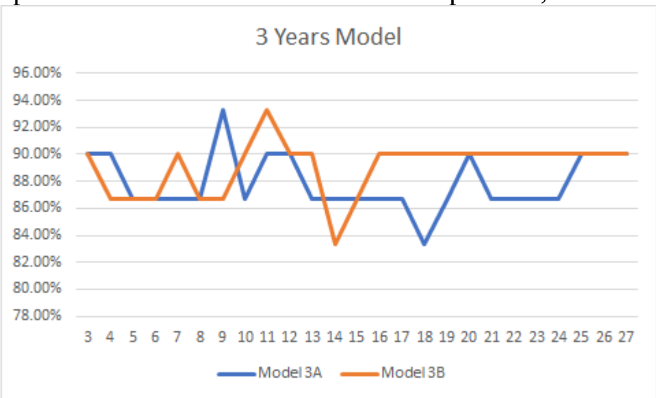

Fig. 4 Accuracy of the 3-year models

*name of corresponding author

The result of the 4-year model accuracy is shown in Figure 4. In this model, we perform feature expansion by testing within a range of 3 to 36 features. The model's highest accuracy value is 90%, achieved with the expansion of 3, 4, 11, 12, 13, 14, 15, 16, 18, 23, 24, 28, 34, and 35 features.

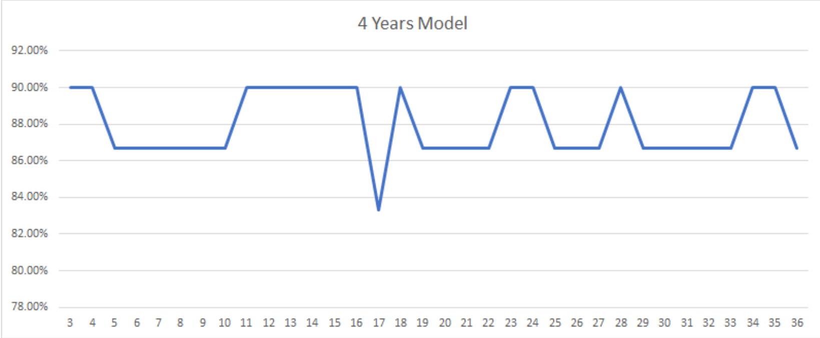

Fig. 5 Accuracy of the 4-year model

DISCUSSIONS

Evaluating accuracy and assessing feature expansion determines the optimal classification model. Although both model 2B with the expansion of 7 and 9 features and model 2C with the expansion of 14 features have an accuracy value of 93.33%, the superior 2-year model is model 2B with the expansion of 9 features. This model includes overall population, proportion of male population, primary school and secondary school graduates, high school diplomas, and university graduates, encompassing the attributes of other models. Although model 2C, which includes the expansion of 14 features, achieves the same level of accuracy as model 2B, it is not chosen as the preferred model. This is because the attributes are already covered in model 2B with the expansion of 9 features.

The same applies to the optimal three-year model, 3B, with the addition of eleven features. The chosen variables include overall population, male population percentage, number of primary, secondary, high school, and university graduates.

Meanwhile, the most optimal 4-year model is model 4A, which expands upon 11 features that include population size, male population percentage, primary school graduates, secondary school graduates, high school diplomas, university graduates, and temperature. This model demonstrates a 90% accuracy rate. However, this may be due to the lack of attribute combinations.

Table 6. The best classification model using CART

Years	Model	Features	Accuracy
2	2B	9	93.33%
3	3B	11	93.33%
4	4A	11	90%

The best classification model using CART with feature expansion is model 3B with expansion of 11 features where the accuracy value is 93.33%, as shown in Table 6. The CART classification model demonstrated superior performance in comparison to prior studies conducted by (Mufid et al., 2018; Shaukat Dar & Ulya Azmeen, 2015). It is because CART is relatively efficient in handling a large number of datasets. However, this CART classification model is lower than (Abdualgalil et al., 2022; Hamdani et al., 2022) because the small number of features used. The study (Abdualgalil et al., 2022) used 22 features and the study (Hamdani et al., 2022) used 18 features. Meanwhile, this study only used 9 features.

CONCLUSION

From the results, it may be concluded that the CART algorithm produced the most accurate classification model over the last two, three, and four years, with 93%, 93%, and 90% accuracy, respectively. Furthermore, feature expansion based on time was found to impact the accuracy of the classification process. Population size, the ratio of males to females, the number of individuals who have completed primary, secondary, and tertiary education are the factors that affect the spread of dengue

*name of corresponding author

hemorrhagic fever (DHF). It is important to understand how these factors interact to develop effective prevention and control measures for DHF. However, meteorological parameters like rainfall, temperature, and humidity were shown to have insignificant influence in the classification process. Additionally, it can be concluded that the number of features has an impact on the accuracy of the classification model. For future research, this study may be utilized to anticipate the transmission of DHF by supplementing attributes and comparing various methods to obtain improved outcomes.

ACKNOWLEDGMENT

The authors gratefully acknowledge the provision of data by Bandung City Meteorology, Climatology, and Geophysics Agency, Bandung City Health Office, and Bandung City Central Statistics Agency, as well as the support of facilities by Telkom University, without which this research would not have been possible.

REFERENCES

Abdualgalil, B., Abraham, S., & M. Ismael, W. (2022). Early Diagnosis for Dengue Disease Prediction Using Efficient Machine Learning Techniques Based on Clinical Data. *Journal of Robotics and Control (JRC)*, *3*(3), 257–268. https://doi.org/10.18196/jrc.v3i3.14387

Arafiyah, R., Hermin, F., Kartika, I. R., Alimuddin, A., & Saraswati, I. (2018). Classification of Dengue Haemorrhagic Fever (DHF) using SVM, naive bayes and random forest. *IOP Conference Series: Materials Science and Engineering*, *434*.

Asadi, F., Trinugroho, J. P., Hidayat, A. A., Rahutomo, R., & Pardamean, B. (2022). Identification of Meteorological Parameters Affecting Dengue Hemorrhagic Fever Incidences. *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 130–135. IEEE. https://doi.org/10.1109/ICIMCIS56303.2022.10017803

Astuti, T., Suciati, Mujiati, I., Ayu, D., Ristianah, V., & Lestari, W. A. (2016). PENERAPAN ALGORITME J48 UNTUK PREDIKSI PENYAKIT DEMAM BERDARAH. *Telematika*.

Center for Epidemiological Data and Surveillance, M. of H. R. of I. (2010). *Dengue Hemorrhagic Fever in Indonesia 1968-2009*. Pusat Data dan Surveilans Epidemiologi Kementrian Kesehatan RI.

Data and Information Center of the Indonesian Ministry of Health. (2018). *Situation of Dengue Fever in Indonesia in 2017*.

Fariza, A., Mu'Arifin, & Astuti, D. W. (2021). Spatial-Temporal Visualization of Dengue Haemorrhagic Fever Vulnerability in Kediri District, Indonesia, Using K-means Algorithm. *2021 International Conference on Data and Software Engineering (ICoDSE)*, 1–6. IEEE. https://doi.org/10.1109/ICoDSE53690.2021.9648487

Ghoneim, S. (2019, April 2). Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?

Gui, H., Gwee, S., Koh, J., & Pang, J. (2021). Weather Factors Associated with Reduced Risk of Dengue Transmission in an Urbanized Tropical City. *Int. J Environ Res Public Health*.

Hamdani, H., Hatta, H. R., Puspitasari, N., Septiarini, A., & Henderi, H. (2022). Dengue classification method using support vector machines and cross-validation techniques. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *11*(3), 1119. https://doi.org/10.11591/ijai.v11.i3.pp1119-1129

Harapan, H., Michie, A., Mudatsir, M., Sasmono, R. T., & Imrie, A. (2019). Epidemiology of dengue hemorrhagic fever in Indonesia: analysis of five decades data from the National Disease Surveillance. *BMC Research Notes*, *12*(1), 350. https://doi.org/10.1186/s13104-019-4379-9

Inayah, F. N., Prasetyowati, S. S., & Sibaroni, Y. (2020). Classification of Dengue Hemorrhagic Fever (DHF) Spread in Bandung Regency using Hybrid Naïve Bayes, K-Nearest Neighbor, and Artificial Neural Network Methods. *Intl. Journal on ICT*, *6*, 1–10.

Irmanita, R., Sri Suryani Prasetiyowati, & Yuliant Sibaroni. (2021). Classification of Malaria Complication Using CART (Classification and Regression Tree) and Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(1), 10–16. https://doi.org/10.29207/resti.v5i1.2770

Ministry of Health. (2020). Latest DHF Case Data in Indonesia.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13023
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Mufid, M. R., Kusuma Saginta Putri, N. R., Fariza, A., & Mu'arifin. (2018). Fuzzy Logic and Exponential Smoothing for Mapping Implementation of Dengue Haemorrhagic Fever in Surabaya. *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, 372–377. IEEE. https://doi.org/10.1109/KCIC.2018.8628533

Shaukat Dar, K., & Ulya Azmeen, S. M. (2015). Dengue Fever Prediction: A Data Mining Problem. *Journal of Data Mining in Genomics & Proteomics*, *06*(03). https://doi.org/10.4172/2153-0602.1000181

Taneja, P., & Gautam, N. (2019). Hybrid Classification Method for Dengue Prediction. *International Journal of Engineering and Advanced Technology (IJEAT)*, *8*(6).

Timofeev, R. (2004). *Classification and Regression Trees (CART) Theory and Applications*. Humboldt University, Berlin.

Wahono, H., & Riana, D. (2020). Prediksi Calon Pendonor Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5. *JURIKOM (Jurnal Riset Komputer)*, *7*, 7–14.

Widodo, S., Brawijaya, H., & Samudi, S. (2022). Stratified K-fold cross validation optimization on machine learning for prediction. *Sinkron*, *7*(4), 2407–2414. https://doi.org/10.33395/sinkron.v7i4.11792

World Health Organization. (2023, March 17). Dengue and severe dengue.

Ye, G., Xu, Z., Yang, M., Wang, J., Liang, J., Yin, J., … Liu, Y. (2023). Clinical features and transmission risk analysis of dengue virus infections in Shenzhen, During 2014–2019. *Computational and Structural Biotechnology Journal*, *21*, 3728–3735. https://doi.org/10.1016/j.csbj.2023.07.001

Zeng, Y., Jiang, K., & Chen, J. (2019). Automatic Seismic Salt Interpretation with Deep Convolutional Neural Networks. *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*, 16–20. New York, NY, USA: ACM. https://doi.org/10.1145/3325917.3325926

*name of corresponding author